THE APPLICATION OF CONVOLUTIONAL CODES TO THE HIGH FREQUENCY CHANNEL

by

James Brayton McCrumb

United States
Naval Postgraduate School

THESIS

THE APPLICATION OF CONVOLUTIONAL CODES
TO THE HIGH FREQUENCY CHANNEL

by

James Brayton McCrumb

September 1970

The Application of Convolutional Codes
to the High Frequency Channel

by

James Brayton McCrumb
Lieutenant, United States Navy
B.A., University of Colorado, 1963

Submitted in partial fulfillment of the
requirements for the degree of

ELECTRICAL ENGINEER

from the

NAVAL POSTGRADUATE SCHOOL
September 1970

ABSTRACT

The HF channel is characterized with regard to frequency selective fading and atmospheric noise under restrictive mathematical assumptions, and compared with measured statistics. Actual channel characteristics are utilized to obtain parameters for choosing appropriate convolutional codes. Short threshold decodable convolutional codes with only a moderate amount of interleaving are shown to be capable of affecting a reduction in the output error probability by at least an order of magnitude, for uncoded error probabilities on the order of 2×10^{-2}.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF CHARACTERS AND ABBREVIATIONS

HF - Frequency portion of the radio spectrum from 3 to 30 MHz

MUF - Maximum usable frequency

LUF - Lowest useful high frequency

$f(x)$ - Probability density function of the continuous random variable x

$E[g(x)]$ - Operation of taking the expected value of $g(x)$

$C(\xi)$ - Expected value of $e^{j\ x}$ where $j = \sqrt{-1}$

J - Jacobian - coordinate transformation from (u,v) to (x,y)

$J_n(x)$ - Bessel function of order n

β - Average probability of an error in digital system

T_s - Signal duration

B - Receiver Bandwidth

σ_n - Variance of zero mean Gaussian noise

$P(V_o)$ - Probability that the envelope of received atmospheric noise exceeds V_o

APD - Amplitude probability distribution

C - Channel capacity

ν - Memory span of convolutional code

N - Actual constraint length of convolutional code

T_i - Column matrix representing the i^{th} set of transmitted encoded digits

M_i - Column matrix representing the i^{th} set of information digits

R_i - Column matrix representing the i^{th} set of received channel digits

S_i - Column matrix representing the i^{th} set of syndromes

P_{DD} - Probability of a decoding error for definite decoding

P_{FD} - Probability of a decoding error for feedback decoding

ACKNOWLEDGEMENT

I am greatly indebted to many people who helped make this research possible. Specifically, to Dr. M. Nesenbergs and the Radio Interference Section of the Institute for Telecommunication Sciences, who provided a wealth of information concerning the HF channel, atmospheric noise and coding theory. In addition, I wish to express my gratitude to my advisor, Dr. George H. Marmont, without whose encouragement, assistance and occasional prodding, this thesis might not have been written.

Blank

I. INTRODUCTION

Radio communication in the frequency range from 2-30 MHz has been common for many years. However, the reliability of such a channel is not high enough for modern high speed computer to computer communications. In order to increase the reliability of this form of communications, the nature of the channel must be understood, or at least modeled in a fashion that is a reasonable approximation to the situation. Once this is accomplished, various solutions may be attempted to obtain optimal reliability via the channel.

To describe the situation in general terms, the transmitter sends a signal through the channel to a receiver. However, due to various causes, the signal received may not be the one sent, in which case an error is said to have occurred. These errors result from two basically different causes: 1) signal distortion, and 2) noise. Noise is considered throughout this report to be completely unrelated to the signal, while signal distortion is considered as a perturbation of the signal and thus is related to the signal. As an example, if a portion of the signal energy is delayed so that it arrives slightly later than the rest of the signal, interference will result. If the interference is destructive, the total received signal will be less and the signal is said to fade.

The first section is devoted to a detailed discussion of the signal and one common form of noise associated with

the channel. The second section discusses convolutional codes in general and applies them to the channel as characterized in the first section. It is shown that it is possible to obtain several orders of magnitude improvement in the probability of not making an error.

II. THE SIGNAL ENVIRONMENT

A. INTRODUCTION

The HF channel is difficult to analyze in general due to the number and complexity of the various parameters. The channel is time varying over a period of hours, days, months, or years. (The eleven year sun spot cycle greatly effects parameters such as Maximum Usable Frequency (MUF), Lowest Useful Frequency (LUF) and absorption.) In general the HF communication channel is not normally line of sight, this implies that there is some form of 'bending' involved with the path of propagation of the electromagnetic radiation. This 'bending' results from two principal sources: 1) refraction caused by a gradient in the dielectric constant of the medium, and 2) reflection caused by an abrupt discontinuity in the dielectric constant. Refraction normally accounts for a considerable amount of the signal energy received and, is commonly referred to as the specular component.

In the way of a basic review, the upper atmosphere becomes ionized in various regions from ultraviolet and corpuscular radiation from the sun [1]. Incident electromagnetic radiation is refracted by these levels and if the operating or carrier frequency is properly chosen the radiated signal will return to earth at the receiver. The amount of energy received is dependent on many factors ranging from antenna polarization to angle of launch or

reception and the exact state of the ionisphere. Frequently multiple paths will be possible and the resulting signals at the receiver may add either constructively or destructively. In addition to the refraction, some of the incident radiation may be scattered from irregularities in the ionisphere and this will also be received. Figures 2.1 and 2.2 illustrate the basic phenomena. Hence the received signal is the sum of the scatter and spectral signals. The probability characteristics of this type of signal have been well treated by Nesenbergs [2] and are included here only for a review.

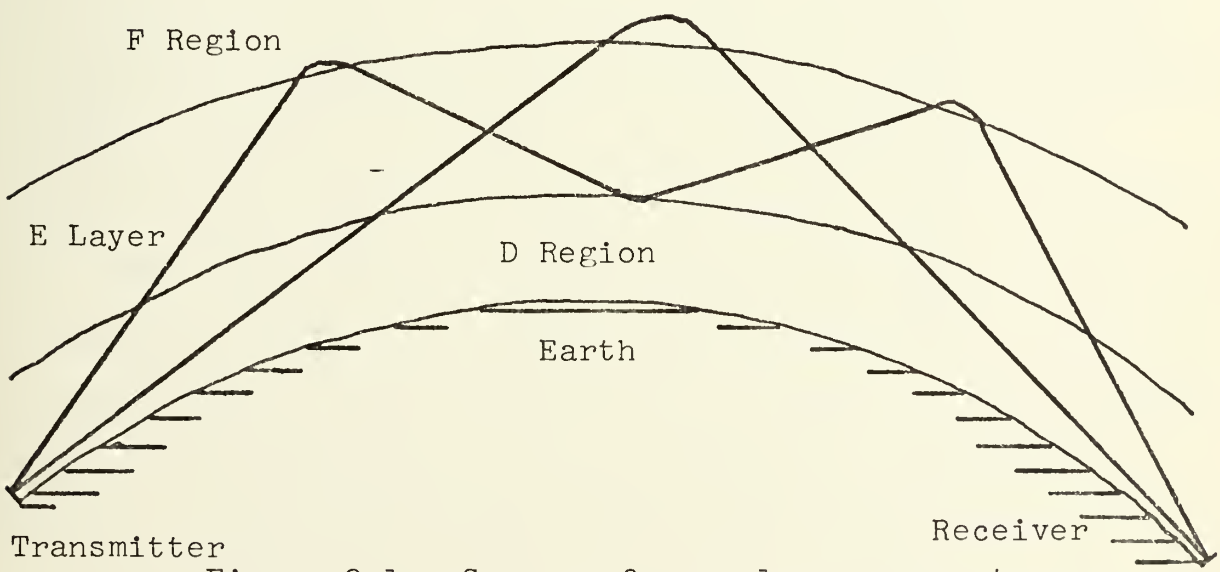

Figure 2.1. Source of specular components. Paths which result in the reception of two specular components at the receiver. The refraction of the incident electromagnetic energy is a result of the gradient of ion density as a function of height.

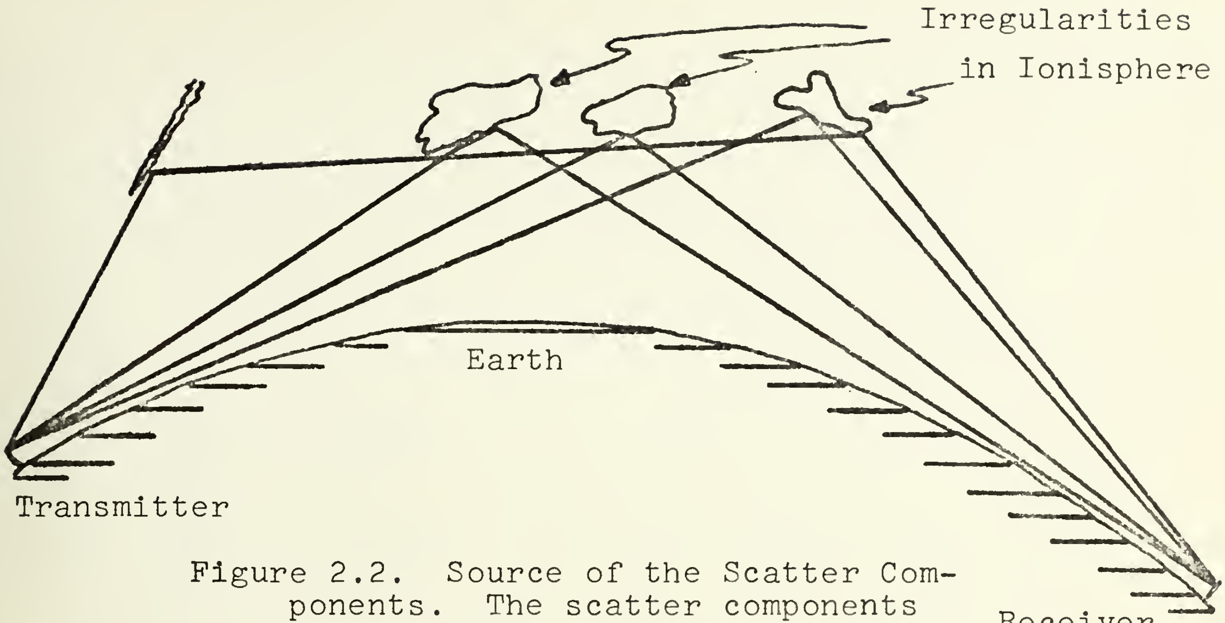

Figure 2.2. Source of the Scatter Components. The scatter components result from the reflection from many relatively abrupt discontinuities in the ionosphere. The discontinuities may be due to numerous causes such as meteor trails or turbulance.

For a given path length, a relatively straight forward computational procedure may be followed to give the maximum usable frequency, and the relative amplitude of other specular components which may be presented with the fraction of time these components may be expected. Hence, for a given operating frequency, a rough approximation may be made with regard to the number of specular components. This, we shall see, will affect the bit error probability. Table I is a set of sample predictions (from Ref. 2) indicating the specular components probable together with the percentage of times they were expected to be present.

It is apparent that this system is quite complicated, so let us begin by making some restrictive assumptions which will facilitate the mathematics.

TABLE I

SAMPLE PREDICTIONS FOR THE MONTH OF DECEMBER 1965, ON THE HAWAII TO STOCKBRIDGE, N.Y., PATH (4200 N. MI.)

Reprinted from Ref. 2 with the kind permission of Dr. Nesenbergs.

G.M. time	Frequency (MHz)	r_1	(%)	r_2	(%)	r_3	(%)	r_4	(%)	r_5	(%)
2:00	7.0	1.00	(45)	0.83	(39)	0.83	(39)	0.73	(34)		
	9.0	1.00	(78)	0.92	(75)	0.87	(73)	0.86	(73)	0.83	(69)
	11.0	1.00	(86)	0.98	(87)	0.96	(79)	0.95	(65)		
	13.0	1.00	(37)	0.97	(83)	0.97	(37)				
	15.0	1.00	(38)	0.97	(67)	0.97	(12)				
	17.0	1.00	(42)	0.99	(14)						
	19.0	1.00	(19)								
4:00	5.0	1.00	(24)	0.73	(17)	0.64	(14)				
	7.0	1.00	(63)	0.88	(58)	0.86	(53)				
	9.0	1.00	(41)	0.99	(75)	0.92	(62)				
	11.0	1.00	(33)	0.98	(60)	0.97	(9)				
	13.0	1.00	(29)	0.98	(7)						
	15.0	1.00	(7)								
6:00	5.0	1.00	(25)	0.81	(20)	0.75	(18)				
	7.0	1.00	(66)	0.96	(51)	0.94	(61)				
	9.0	1.00	(46)	0.99	(19)	0.96	(71)				
	11.0	1.00	(38)	0.98	(12)						
	13.0	1.00	(24)	0.86	(10)						
	15.0	1.00	(5)								
8:00	5.0	1.00	(27)	0.83	(22)	0.78	(21)				
	7.0	1.00	(68)	0.98	(49)	0.94	(62)				
	9.0	1.00	(39)	0.96	(14)	0.94	(68)				
	11.0	1.00	(53)	0.92	(31)	0.84	(8)				
	13.0	1.00	(18)	0.98	(6)						

TABLE I (Continued)

G.M. time	Frequency (MHz)	r_1	(%)	r_2	(%)	r_3	(%)	r_4	(%)	r_5	(%)
10.00	5.0	1.00	(24)	0.83	(18)	0.79	(16)				
	7.0	1.00	(68)	0.93	(53)	0.92	(62)				
	9.0	1.00	(50)	1.00	(19)	0.97	(73)				
	11.0	1.00	(41)	0.98	(11)						
	13.0	1.00	(24)	0.85	(8)						
12:00	5.0	1.00	(22)	0.78	(16)	0.72	(14)				
	7.0	1.00	(67)	0.94	(49)	0.92	(60)				
	9.0	1.00	(43)	0.97	(14)	0.94	(69)				
	11.0	1.00	(33)	0.98	(7)						
	13.0	1.00	(5)								
14:00	5.0	1.00	(21)	0.94	(20)	0.68	(14)				
	7.0	1.00	(63)	0.83	(58)	0.78	(56)	0.77	(56)	0.75	(55)
	9.0	1.00	(77)	0.97	(77)	0.96	(58)	0.94	(74)		
	11.0	1.00	(50)	0.99	(9)	0.98	(78)				
	13.0	1.00	(45)	0.98	(6)						
	15.0	1.00	(6)								
16:00	7.0	1.00	(13)	0.79	(10)	0.72	(8)	0.56	(6)		
	9.0	1.00	(54)	0.91	(48)	0.88	(47)	0.87	(46)	0.69	(43)
	11.0	1.00	(75)	0.98	(74)	0.89	(71)	0.88	(63)		
	13.0	1.00	(35)	0.96	(84)	0.92	(71)				
	15.0	1.00	(37)	0.97	(76)						
	17.0	1.00	(44)	0.98	(5)						
	19.0	1.00	(10)								

TABLE I (Continued)

G.M. time	Frequency (MHz)	r_1	(%)	r_2	(%)	r_3	(%)	r_4	(%)	r_5	(%)
18:00	11.0	1.00	(21)	0.92	(19)	0.69	(13)	0.61	(13)		
	13.0	1.00	(53)	0.72	(43)	0.62	(38)				
	15.0	1.00	(70)	0.83	(64)	0.79	(60)				
	17.0	1.00	(80)	0.92	(58)	0.91	(76)				
	19.0	1.00	(26)	0.97	(86)	0.94	(71)				
	21.0	1.00	(45)	0.96	(81)						
	23.0	1.00	(62)	1.00	(12)						
20:00	11.0	1.00	(5)	0.60	(1)	0.59	(1)				
	13.0	1.00	(38)	0.67	(23)	0.63	(22)	0.51	(16)	0.49	(15)
	15.0	1.00	(59)	0.79	(50)	0.72	(45)				
	17.0	1.00	(76)	0.88	(56)	0.87	(71)				
	19.0	1.00	(30)	0.98	(86)	0.92	(73)				
	21.0	1.00	(50)	0.97	(85)						
	23.0	1.00	(68)	1.00	(15)						
22:00	11.0	1.00	(19)	0.59	(9)	0.50	(7)	0.45	(5)	0.35	(3)
	13.0	1.00	(43)	0.74	(33)	0.70	(30)	0.66	(29)	0.60	(26)
	15.0	1.00	(64)	0.85	(58)	0.82	(55)	0.81	(57)		
	17.0	1.00	(78)	0.97	(52)	0.92	(74)				
	19.0	1.00	(16)	0.97	(86)	0.96	(68)				
	21.0	1.00	(33)	0.96	(82)						
	23.0	1.00	(56)	0.98	(6)						
24:00	9.0	1.00	(22)	0.60	(11)	0.53	(9)	0.47	(7)	0.39	(5)
	11.0	1.00	(50)	0.79	(41)	0.78	(40)	0.75	(39)	0.72	(38)
	13.0	1.00	(72)	0.92	(69)	0.88	(68)	0.86	(65)		
	15.0	1.00	(53)	0.99	(84)	0.94	(80)				
	17.0	1.00	(11)	0.99	(64)	0.97	(89)				
	19.0	1.00	(20)	0.97	(76)						
	21.0	1.00	(38)								
	23.0	1.00	(8)								

We assume that the HF channel under consideration is stationary. While this is not strictly true, over the period of several minutes it is a very reasonable approximation. A second helpful assumption is that all the components (scatter and specular) are mutually statistically independent. This assumption will permit us to use characteristic functions and simplify our task. (The number of modes possible may be determined by the relationship of the carrier frequency with the maximum usable frequency, however, if the modes exist their effects may be statistically independent.)

It will also be assumed in this section that the environmental noise is gaussian. In a later section we will consider in some detail a recent model for atmospheric noise and indicate the modifications which result.

B. SIGNAL AMPLITUDE CHARACTERISTICS

We begin by considering the received signal as two dimensional, where the two dimensions will be the inphase and quadrature components of the signal. We will have occasion to consider two basic types of signals which may be present at the same time, hence we define a signal to be specular if it is of fixed amplitude, although the phase may still be random. A second type of signal, called scatter, is characterized by a symmetric bivariate gaussian distribution which for mathematical convenience has equal inphase and quadrature variance. We define the probability distribution function (DF) of the random variable X as

$$F(x) = \text{Prob}\ (X \leq x) \qquad -\infty < x < \infty$$

with the corresponding probability density function (pdf)

$$f(x) = \frac{dF(x)}{dx} .$$

It will be convenient to define, in the standard fashion, the characteristic function

$$\begin{aligned} C(\xi) &= E\left[e^{j\xi x}\right] \\ &= \int_{-\infty}^{\infty} e^{j\xi x}\, dF(x) \\ &= \int_{-\infty}^{\infty} e^{j\xi x}\, f(x)dx . \end{aligned}$$

where $E[\cdot]$ denotes the expected value.

For our purpose, we consider the two dimensional case where

$$F(x,y) = \text{Prob}(X \leq x,\ Y \leq y)$$

with the corresponding joint probability density function $f(x,y)$ and characteristic function

$$C(\xi,\eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j(\xi x+\eta y)} f(x,y)dxdy . \tag{1}$$

Let us consider a typical specular component,

$$V_i = X_i + jY_i = |V_i|\ e^{j\phi_i}$$

where X_i is the cosine component, Y_i the sine component, $|V_i|$ is the envelope and ϕ the phase of V_i. Let us assume that the phase is uniformly distributed over $(0,2\pi)$, and independent of $|V_i|$. Since we are dealing with a specular

component, which, by definition, has a constant amplitude, we have

$$f_{|V_i|}(x,y) = \delta\left(\sqrt{x_i^2 + y_i^2} - r_i\right),$$

where x_j and y_i are the inphase and quadrature components respectively and $r_i = \sqrt{x_i^2 + y_i^2}$

$$f_i(\phi_i) = \frac{1}{2\pi}$$

and since $|V_i|$ and ϕ_i are independent

$$f_{|V_i|,\phi_i}(\rho,\phi) = f_{|V_i|}(\rho)\cdot f_{\phi_i}(\phi) .$$

We desire the pdf of V_i, which may be expressed as

$$\iint_{V\Phi} f_{|V|,\phi}(\rho,\phi)\, d\rho d\phi = \iint_{XY} f_{V_i}(x,y)\, dxdy$$

Changing variables to $x = \rho\cos\phi$, $y = \rho\sin\phi$, we have the Jacobian

$$|J| = |\rho|$$

so we now have

$$\iint_{\rho\Phi} f_{|V|,\phi}(\rho,\phi)\, d\rho d\phi = \iint_{\rho\Phi} \rho f_{V_i}(\rho\cos\phi,\ \rho\sin\phi)\, d\rho d\phi$$

from whence

$$f_{|V|,\phi}(\rho,\phi) = \rho f_{V_i}(\rho\cos\phi,\ \rho\sin\phi)$$

$$= \rho f_{V_i}(x,y)$$

and

$$f_{V_i}(x,y) = \frac{1}{2\pi\sqrt{x^2+y^2}}\ f_{|V|}\left(\sqrt{x^2+y^2}\right) \qquad (2)$$

From equation 1 we may write the characteristic function for the specular component as

$$C_i(\xi,\eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{j(\xi x+\eta y)}\ \delta\left(\sqrt{x^2+y^2} - r_i\right)\ dxdy}{2\pi\sqrt{x^2+y^2}}$$

Again letting $x = \rho\cos\phi$, $y = \rho\sin\phi$; the Jacobian is again ρ and

$$C_i(\xi,\eta) = \int_0^{2\pi} \int_0^{\infty} \frac{e^{j\rho(\xi\cos\phi+\eta\sin\phi)}\delta(\rho-r_i)}{2\pi\rho} \cdot |\rho|\, d\rho d\phi$$

The exponent may be reduced by noting:

$$\begin{aligned} \xi\cos\phi + \eta\sin\phi &= \sqrt{\xi^2+\eta^2}\left(\frac{\xi}{\sqrt{\xi^2+\eta^2}}\cos\phi + \frac{\eta}{\sqrt{\xi^2+\eta^2}}\sin\phi\right) \\ &= \sqrt{\xi^2+\eta^2}\cos(\phi+\alpha) \\ &= \zeta\cos\phi' \end{aligned}$$

where $\zeta = \sqrt{\xi^2+\eta^2}$ and ϕ' is again uniform on $(0,2\pi)$, hence,

$$C_i(\xi,\eta) = C_i(\zeta) = \int_0^{\infty} \delta(\rho-r_i)\cdot \frac{1}{\pi}\int_0^{\pi} e^{j\zeta\rho\cos\phi'} d\phi' d\rho$$

From Reference 3, item 9.1.21

$$J_n(x) = \frac{(j)^{-n}}{\pi}\int_0^{\pi} e^{jx\cos\theta}\cos(n\theta)d\theta$$

hence

$$C_i(\zeta) = \int_0^{\infty} \delta(\pi-r_i)\ J_0(\zeta\rho)d\rho \qquad (3)$$

$$= J_0(\xi r_i)$$

where $\zeta = \sqrt{\xi^2+\eta^2}$.

Turning our attention to one of the scatter components, which is characterized by the bivariate gaussian distribution, we have the pdf:

$$f_{S_i}(x,y) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{x^2}{\sigma_i^2/2} + \frac{y^2}{\sigma_i^2/2}}$$

$$= \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{2\rho^2}{\sigma_i^2}}$$

where again $\rho = \sqrt{x^2+y^2}$.

From Reference 3, item 26.1.26 the characteristic function is

$$c_{S_i}(\zeta) = e^{-\frac{\sigma_j^2\zeta^2}{4}} \qquad (5)$$

where again $\zeta = \sqrt{\xi^2+\eta^2}$.

We now define the scatter vector to be the sum of all the independent scatter components. Our motive for proceeding to the characteristic functions becomes clear since the characteristic function of a sum of independent random vectors is simply the product of their characteristic functions. Hence, we have for the characteristic function of the scatter component:

$$C_S(\zeta) = \prod_{i=1}^{n} e^{-\frac{\sigma_i^2 \zeta^2}{4}}$$

$$= e^{-\frac{\zeta^2}{4} \sum_{i=1}^{n} \sigma_i^2} \tag{6}$$

$$= e^{-\frac{\sigma^2 \zeta^2}{4}} .$$

where $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$. The characteristic function of this scatter vector and one specular component is then

$$C(\zeta) = C_S(\zeta)\, C_i(\zeta)$$

$$= e^{-\frac{\sigma^2 \zeta^2}{4}} J_o(\zeta r_i)$$

By adding a total of n specular components the characteristic function becomes:

$$C(\zeta) = e^{-\frac{\sigma^2 \zeta^2}{4}} \prod_{i=1}^{n} J_o(\zeta r_i) \tag{7}$$

The characteristic function, as we have developed it here, is convenient for expressing only the even moments since

$$E[\rho^{2m}] = (-1)^m \left(\frac{\partial^2}{\partial \zeta^2} + \frac{1}{\zeta} \frac{\partial}{\partial \zeta} \right)^m C(\zeta) \Big|_{\zeta=0} .$$

The second moment (m = 1) is:

$$E[\rho^2] = \sigma^2 + \sum_{i=1}^{n} r_i^2$$

The pdf for $|V|$ may be expressed in terms of the Humbert hypergeometric series (see Ref. 2):

$$f_{|V|}(\rho) = \frac{2\rho}{\sigma^2} \psi_2 \left(1;1,1,\ldots,1; -\frac{\rho 2}{\sigma^2}, -\frac{r_1^2}{\sigma^2}, \ldots \frac{r_n^2}{\sigma^2} \right) \quad (8)$$

Let us now restrict ourselves to digital systems where we may relate the signal pdf above, to the probability of an error β. We define the probability of an error as the expected value, averaged over the distribution of the signal amplitude, of a "probability of error function." In other words

$$\begin{aligned} \beta &= E[p_e(\rho)] \\ &= \int_0^\infty p_e(\rho)\, f_{|V|}(\rho) d\rho \end{aligned} \quad (9)$$

The 'probability of error function' is determined by the modulation/detection procedure and the noise environment. It represents the probability of making an error at a particular time as a function of the signal amplitude and a given type of noise. Equivalently, it may be viewed as either the probability of a false alarm or the probability of a miss in detection theory. We consider only the case of noncoherent frequency-shift-keying in stationary, zero mean, gaussian noise. If the signal duration is T_s, the receiver bandwidth B, and the total noise power is the receiver pass band is N_o, then the probability of making an error for a given signal amplitude is given by Montgomery [4] as:

$$p_e(\rho) = \tfrac{1}{2} e^{-\frac{T_s B \rho^2}{2N_0}} . \tag{10}$$

We note, in passing, that in case of constant signal energy where $\rho = \sqrt{E}$, our definition of the probability of an error is consistent; since in this case,

$$f_V(\rho) = \delta(\rho - \sqrt{E})$$

and

$$\beta = \int_0^\infty \tfrac{1}{2} e^{-\frac{T_S B \rho^2}{2N_0}} \delta(\rho - \sqrt{E})\, d\rho$$

$$= \tfrac{1}{2} e^{-\frac{T_S B E}{2N_0}}$$

which is the well known result, giving the bit error probability as a function of signal energy/noise energy.

With the signal duration, T_s, receiver bandwidth, B, and noise power N_0, we may utilize equations 9 and 10 together with the envelope pdf given in equation 8 to obtain the probability of an error. The result, from Nesenbergs [Ref. 2], is

$$\beta = \frac{N_0}{T_s B} \cdot \frac{1}{\sigma^2 + \frac{2N_0}{T_s B}} \psi_2 \left(1;1,1,\ldots,1; \frac{r_1^2}{\sigma^2 + \frac{2N_0}{T_s B}}, \ldots, \frac{r_n^2}{\sigma^2 + \frac{2N_0}{T_s B}} \right) \tag{11}$$

where r_i = amplitude of the i^{th} specular component and σ^2 variance of the scatter components of the received signal.

Nesenbergs has developed bounds for this hypergeometric series in Reference 2 which yield:

$$\frac{1}{2}\,\frac{e^{-\frac{r^2}{\sigma^2+\frac{2N_0}{T_sB}}}}{1+\frac{T_sB\sigma^2}{N_o}} \leq \beta \leq \frac{1}{2}\,\frac{1}{1+\frac{T_sB(r^2+\sigma^2)}{2N_0}}\left[1-\left(\frac{\frac{T_sBr^2}{2N_0}}{1+\frac{T_sB(r^2+\sigma^2)}{2N_0}}\right)^2\right]^{\frac{1}{2}}$$

where $r^2 = \Sigma\, r_i^2$. Figure 2.3 is a plot of some special cases calculated by Nesenbergs in Ref. 2.

It can be seen from Table I that as the frequency is decreased below the MUF, other propagation paths may occur, resulting in the bit error probability as described in Equation 11. While for frequencies well below the MUF, the signal amplitude pdf may be very nearly Rayleigh, for frequencies near the MUF the Rayleigh approximation is not appropriate. One could reasonably expect the ratio of the operating frequency to the MUF will be roughly related to the bit error probability. Increasing the ratio toward unity results in fewer modes with a corresponding decrease in the error probability, while a decrease in the ratio indicates the possible existance of more modes causing an increase in the error probability. This effect is experienced during periods of the day, such as sunrise or sunset, when the MUF is changing rapidly, and the carrier frequency is fixed. Tsai [5] and McManamon [6] have also noted this effect in more quantitative terms.

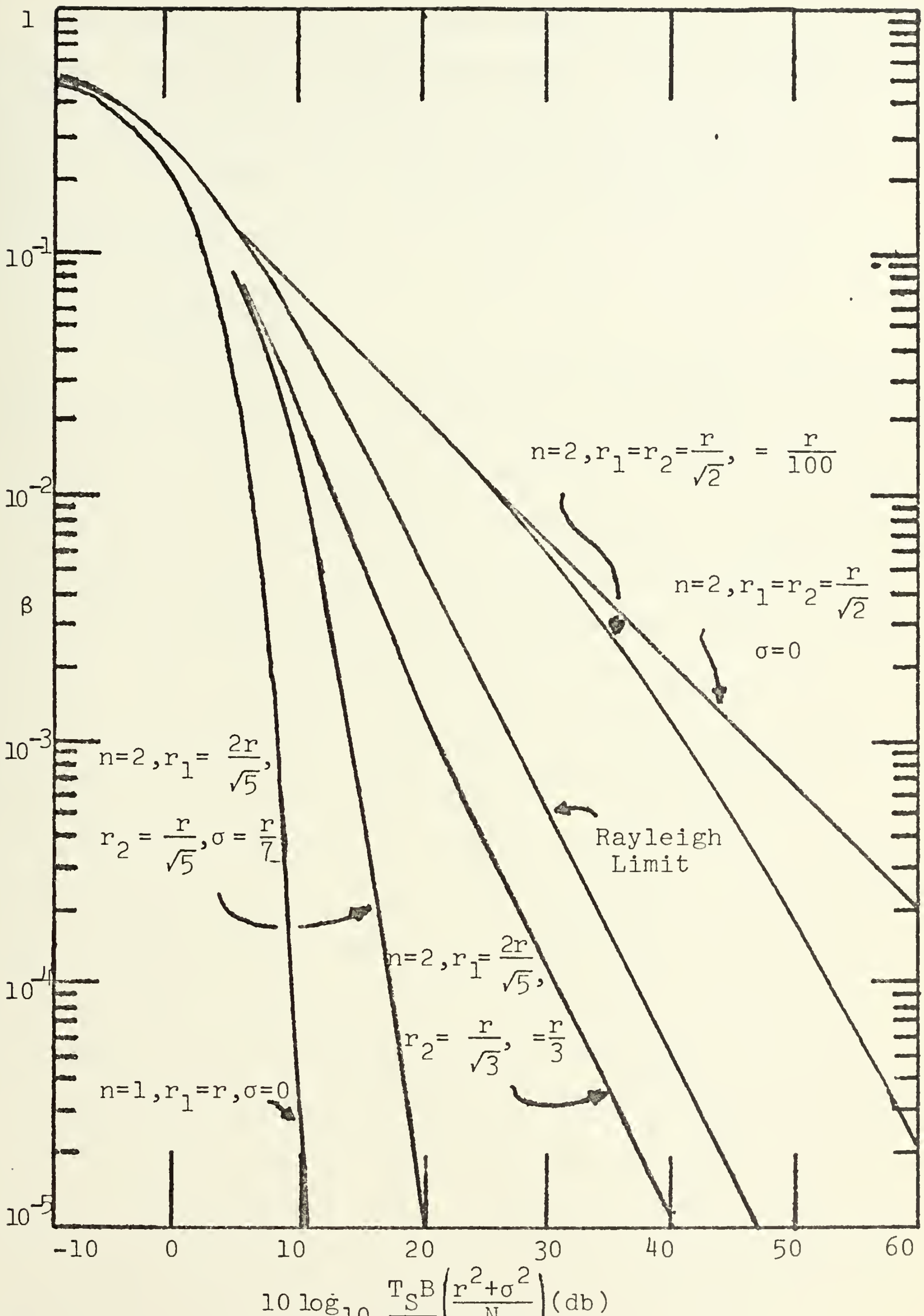

Figure 2.3. Bit error probability β as a function of signal to noise ratio (db). (Reprinted from Ref. 2, p. 82, with permission.)

There are, of course, other sources of error in the HF channel. One such source, which may have an appreciable effect on the error probability, is atmospheric noise. The next section deals with this noise in some detail.

C. ATMOSPHERIC NOISE

1. Introduction

Atmospheric noise is the name generally associated with electromagnetic energy which has its source in natural phenomena such as lightning discharges. In general, atmospheric noise is impulsive and nonstationary. Geographically, the largest dynamic range occurs within the tropics, from whence it propagates into the northern and southern hemispheres. For any particular location the received rms or average power will be determined by local thunderstorm activity, as well as the propagation conditions to the tropics. Although the largest dynamic range (on the order of 100 db) occurs in the lower portion of the frequency spectrum (ELF-LF), it may be the limiting type of noise within a major portion of the HF spectrum. Statistical parameters such as average power, and average envelope voltage have been measured quite accurately and the results published in the form of Technical Notes by the National Bureau of Standards, and the International Radio Consultative Committee (CCIR Report 322, "World Distribution and Characteristics of Atmospheric Radio Noise"[7]).

Numerous models for atmospheric noise are contained in the current literature [Refs. 8 and 9]. Most of these models are based on an impulse model with the impulses distributed according to some particular rule. Models such as these have two principal difficulties, first the parameters of the distribution are chosen so they match the first order statistics of the received noise and have very little correlation with statistics such as pulse interval distribution or pulse duration distribution. Secondly, there may be serious computational difficulties. A more recent model was proposed by H. M. Hall [Ref. 10]. The Hall model seems to fit the measured statistics of the noise and hence will be used for the remainder of this section.

2. The Hall Model

The noise y(t) is considered to be the product of a zero mean narrow bandpass gaussian process n(t) with a covariance function $R_n(\tau)$, and a slowly varying regime process A(t) which is assumed to be independent of n(t) and stationary. Although the actual atmospheric noise is non-stationary, it may be approximated by a stationary process for a reasonably short period of time. The statistics of the modulating regime process are chosen so the resulting product $y(t) = A(t) \cdot n(t)$, is an accurate description of the received noise.

Summarizing Hall's treatment we have the following pdf for the regime process

$$f_a(a) = k \frac{1}{|a|^{m+1}} e^{-\left[\frac{m}{2\sigma^2} \cdot \frac{1}{a^2}\right]} \qquad -\delta \leq a \leq \delta \qquad (12)$$

(a chi distribution with parameters m and σ), where k is chosen so that

$$\int_{-\delta}^{\delta} f_a(a)da = 1 \qquad (13)$$

The use of a finite limit for the amplitude is based on the required convergence of the integral for $1 \leq m \leq 2$, since the energy radiated must be finite, and to match the observed levels at the higher frequencies in the HF band.

The gaussian component is

$$f_n(n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{n^2}{2\sigma_n^2}} . \qquad (14)$$

The calculation of the pdf for y(t) is straight forward [Ref. 11]

$$f_y(y) = \int_{-\infty}^{\infty} p_a(x) \; P_n \frac{y}{x} \; \frac{dx}{|x|} \qquad (15)$$

Substituting equations 13 and 41 into 15 results in

$$f_y(y) = \frac{k}{\sqrt{2\pi\sigma_n^2}} \int_{-\infty}^{\infty} \frac{1}{|x|^{m+2}} e^{-\left(\frac{m}{2\sigma^2} + \frac{y^2}{2\sigma_n^2} \; \frac{1}{x^2}\right)} dx$$

Noting that the integrand is an even function and making the substitution $\xi = \frac{1}{x^2}$, $f_y(y)$ becomes

$$f_y(y) = \frac{k}{\sqrt{2\pi\sigma_n^2}} \int_{\frac{1}{\delta^2}}^{\infty} \xi^{\frac{m-1}{2}} e^{-\left(\frac{m}{2\sigma^2} + \frac{y^2}{2\sigma_n^2}\right)\xi} d\xi$$

or equivalently

$$f_y(y) = \frac{k}{\sqrt{2\pi\sigma_n^2}} \frac{1}{\left(\frac{m}{2\sigma^2} + \frac{y^2}{2\sigma_n^2}\right)^{\frac{m+1}{2}}} \int_{\frac{1}{\delta^2}\left(\frac{m}{2\sigma^2} + \frac{y^2}{2\sigma_n^2}\right)}^{\infty} \lambda^{\frac{m-1}{2}} e^{-\lambda} d\lambda \qquad (16)$$

If we consider the case where $\delta \to \infty$, then

$$k = \frac{\left(\frac{m}{2}\right)^{m/2}}{\sigma^m \, \Gamma\left(\frac{m}{2}\right)}$$

and equation 16 reduces to

$$f_y(y) = \frac{\left(\frac{m^{\frac{1}{2}}\sigma_n}{\sigma}\right)^m \Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi}\, \Gamma\left(\frac{m}{2}\right)\left(\frac{m\sigma_n^2}{\sigma^2} + y^2\right)^{\frac{m+1}{2}}} \qquad (17)$$

In order to compare this mathematical model with the measured atmospheric noise we must find the complement of the probability distribution function for the envelope and phase. We let

$$y(t) = V(t)\text{Cos}(\omega_o t + \phi).$$

The pdf of the envelope and phase may then be obtained in terms of the inphase and quadrature components of the received noise as

$$\iint\limits_{V\Phi} f_{r,\phi}(V,\phi)d\phi dV = \iint\limits_{Y\tilde{Y}} f_{y,\tilde{y}}(y,\tilde{y})dyd\tilde{y}$$

where $f_{y,\tilde{y}}(y,\tilde{y})$ is the joint pdf of the inphase (y) and quadrature $(\tilde{y})$ component of the received noise. Changing variables on the right hand side of the equality sign to $y = V\text{Cos}\phi$, $\tilde{y} = V\text{Sin}\phi$, yields the Jacobian, $J = V(\text{Cos}^2\phi + \text{Sin}^2\phi) = V$ and

$$\iint\limits_{V\Phi} f_{V,\phi}(V,\phi)d\phi dV = \iint\limits_{V\Phi} V\, f_{y,\tilde{y}}(V\cos\phi,\ V\sin\phi)\, dVd\phi$$

hence

$$f_{r,\phi}(V,\phi) = V\, f_{y,\tilde{y}}(V\cos\phi,\ V\sin\phi)$$

We let $\tilde{n}(t)$ represent the quadrature component of the narrow band gaussian noise and $y(t) = A(t)n(t)$ and $\tilde{y}(t) = A(t)\tilde{n}(t)$ where $f_{\tilde{n}}(\tilde{n}) = f_n(n) = N[0,R_n(0)]$, and $n(t)$ is independent of $\tilde{n}(t)$. We may now express the joint pdf as:

$$f_{y,\tilde{y}}(y,\tilde{y}) = \int_{-\infty}^{\infty} f_a(x)\, f_n\left(\frac{y}{x}\right) f_{\tilde{n}}\left(\frac{\tilde{y}}{x}\right) \frac{1}{x^2}\, dx$$

$$= \frac{k}{2\pi\sigma_n^2} \cdot \frac{1}{\left(\dfrac{m}{2\sigma^2} + \dfrac{y^2+\tilde{y}^2}{2\sigma_n^2}\right)^{\frac{m+2}{2}}} \int_{\frac{1}{\delta^2}\left(\frac{m}{2\sigma^2}+\frac{y^2+\tilde{y}^2}{2\sigma_n^2}\right)}^{\infty} \lambda^{\frac{m}{2}} e^{-\lambda} d\lambda$$

and

$$f_{V,\phi}(V,\phi) = \frac{1}{2\pi} \cdot \frac{k\ V}{\sigma_n^2\left(\dfrac{m}{2\sigma^2} + \dfrac{V^2}{2\sigma_n^2}\right)^{\frac{m+2}{2}}} \int_{\frac{1}{\delta^2}\left(\frac{m}{2\sigma^2} + \frac{V^2}{2\sigma_n^2}\right)}^{\infty} \lambda^{\frac{m}{2}}\ e^{-\lambda} d\lambda \tag{18}$$

If again we let $\delta \to \infty$, then

$$f_{V,\phi}(V,\phi) = \frac{V\ m(\gamma)^m}{2\pi\left(\gamma^2+V^2\right)^{\frac{m}{2}+1}} = f_\phi(\phi)\cdot f_V(V) \tag{19}$$

where

$$f_V(V) = \frac{m\gamma^m V}{\left(\gamma^2+V^2\right)^{\frac{m}{2}+1}}\ ,\quad f_\phi(\phi) = \frac{1}{2\pi}\ ,\ \text{and}\ \gamma^2 = \frac{m\sigma_n^{\ 2}}{\sigma^2}\ .$$

The complement of the desired distribution function is simply the probability that the noise envelope voltage exceeds some threshold, V_0, i.e.

$$P(V > V_0) = \int_{V_0}^{\infty} p_V(V)dV = \frac{\gamma^m}{\left(V_0^2 + \gamma^2\right)^{\frac{m}{2}}}$$

For convenience we define

$$P(V_0) = P(V > V_0) = \frac{1}{\left(\left(\frac{V_0}{\gamma}\right)^2 + 1\right)^{\frac{m}{2}}} \tag{20}$$

Figure 2.4 is a plot of the envelope level (in db above γ) vs. $-\log_{10}|\ln P(V_0)|$ for various values of m. Figure 2.5 and 2.6 are plots of actual measured atmospheric noise at

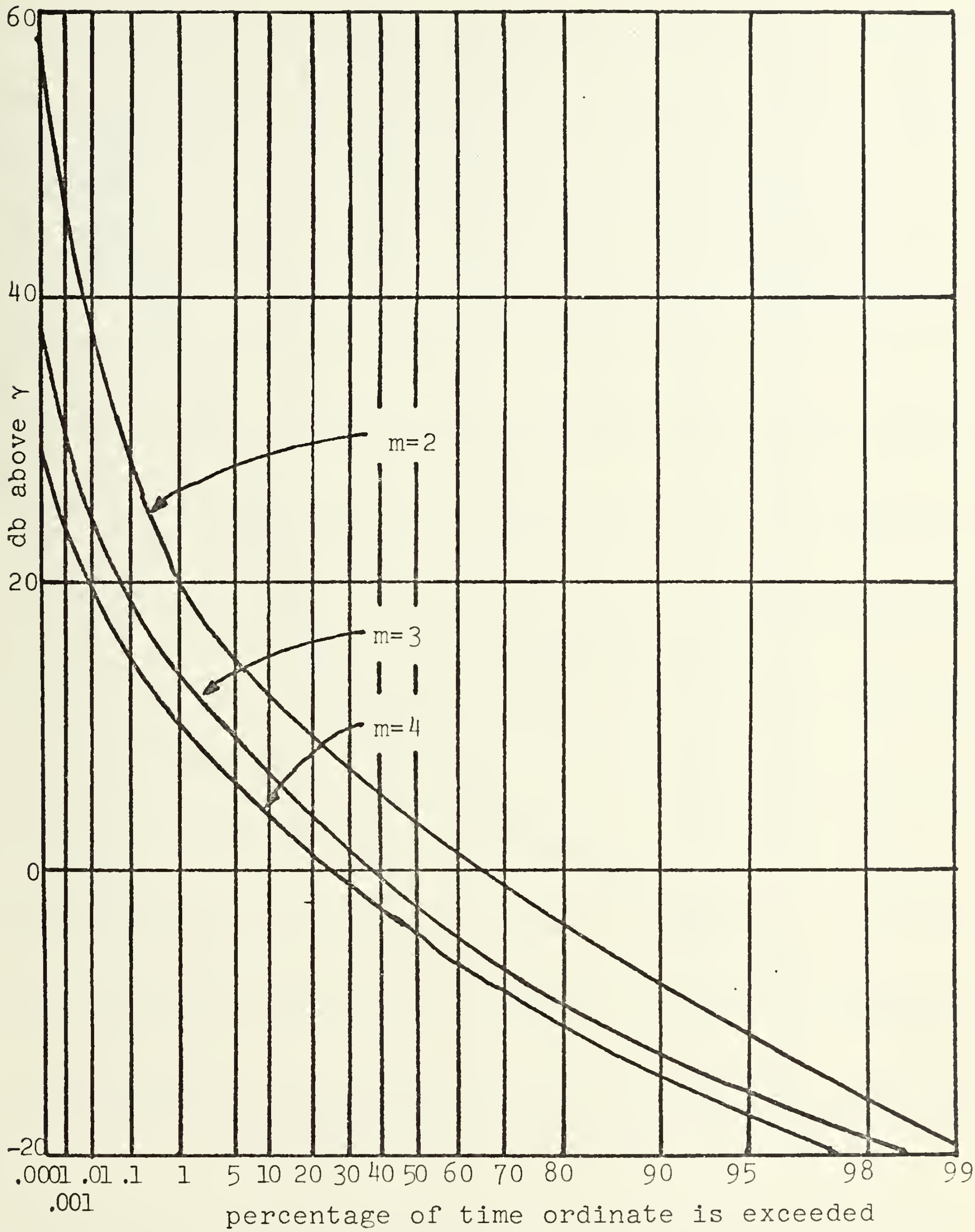

Figure 2.4. APD's obtained from the Hall model for various values of m

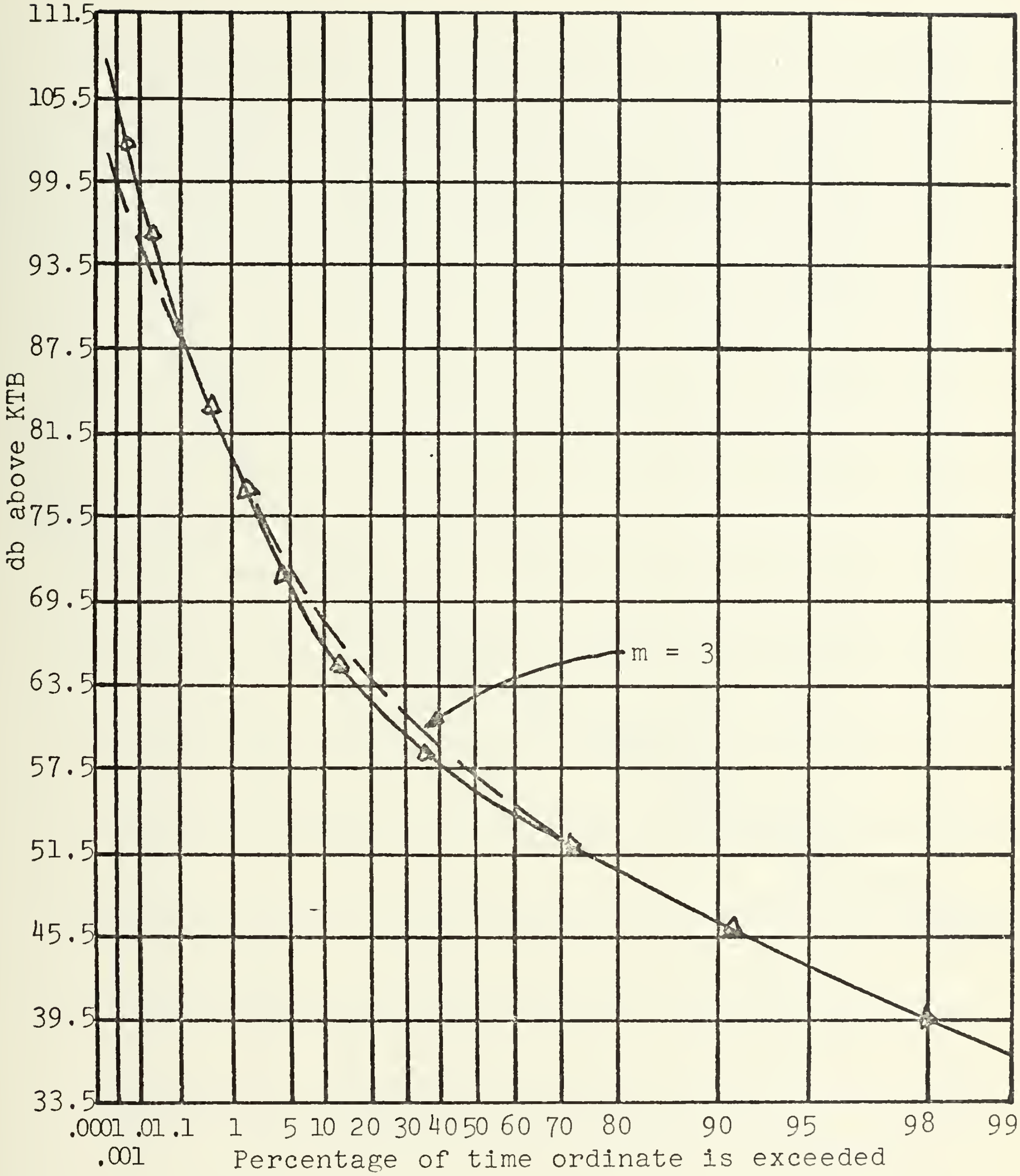

Figure 2.5. APD of atmospheric noise measured at Laramie, Wyo. 2350-2358 MST 16FEB 1970, at a frequency of 2.5 MHz. Δ indicates measured values, the dashed line is obtained from the Hall model with m = 3.

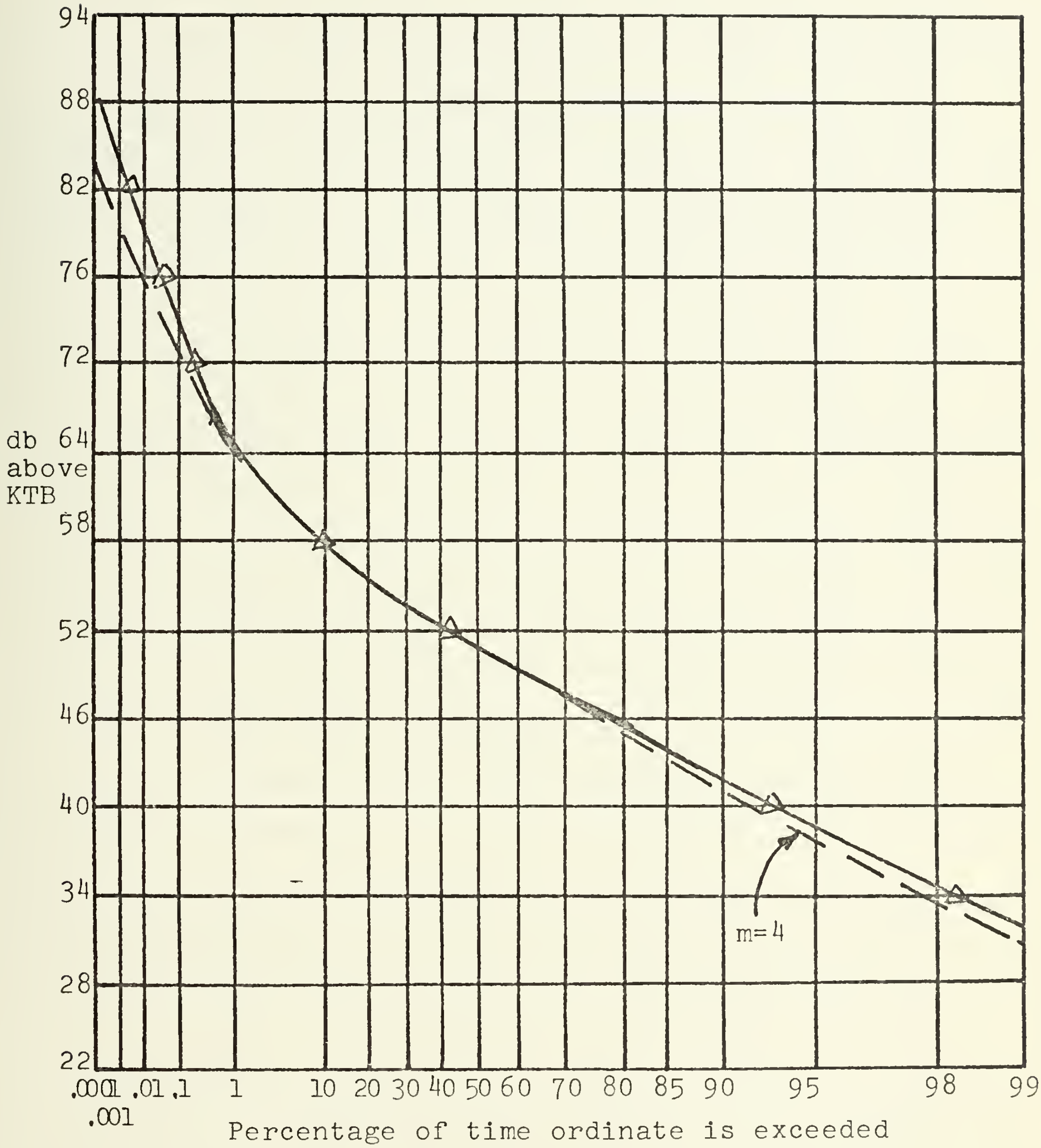

Figure 2.6. APD of atmospheric noise measured at Laramie, Wyo. 2335-2343 MST 17 FEB 1970, at a frequency of 5.0 MHz. Δ indicates measured values, the dashed line is obtained from the Hall model with m = 4.

frequencies of 2.5 MHz and 5.0 MHz respectively.* A plot of the form of figure is an amplitude probability distribution (APD) and represents the percentage of time a given level is exceeded as a function of the level. The plotting paper is designed so that thermal, or gaussian noise plots as a straight line with slope $-\frac{1}{2}$. Montgomery, in 1953, [4] showed that, for FM or FSK modulation systems, the probability of a bit error is one half the probability that the noise amplitude exceeds the signal amplitude. Hence if the signal-to-noise ratio is 10 db, one has simply to find the percentage of time the level corresponding to 10 db above the rms noise level is exceeded. This percentage will then be twice the bit error probability. As an aside, an approximation to the APD may be inferred from the data in CCIR Report 322 [7].

As in the case of the signal amplitude, we have only obtained an 'average' bit error probability. There is no indication of the correlation between errors. Present research (Spaulding, Disney, and Espeland [12]) indicates that the burst interval distribution and burst duration distribution may be matched as closely as desired by proper specification of the covariance function for the modulating regime process A(t). While this is an important factor, it is not within the intent of the present paper, so we will simply give some typical cases to indicate the characteristics of the burst parameters. Figures 2.7 and 2.8 are

---

*The data presented here was obtained at Laramie, Wyoming during the period 16-18 Feb., 1970 while the author was in guest worker status with the Environmental Science Services Administration.

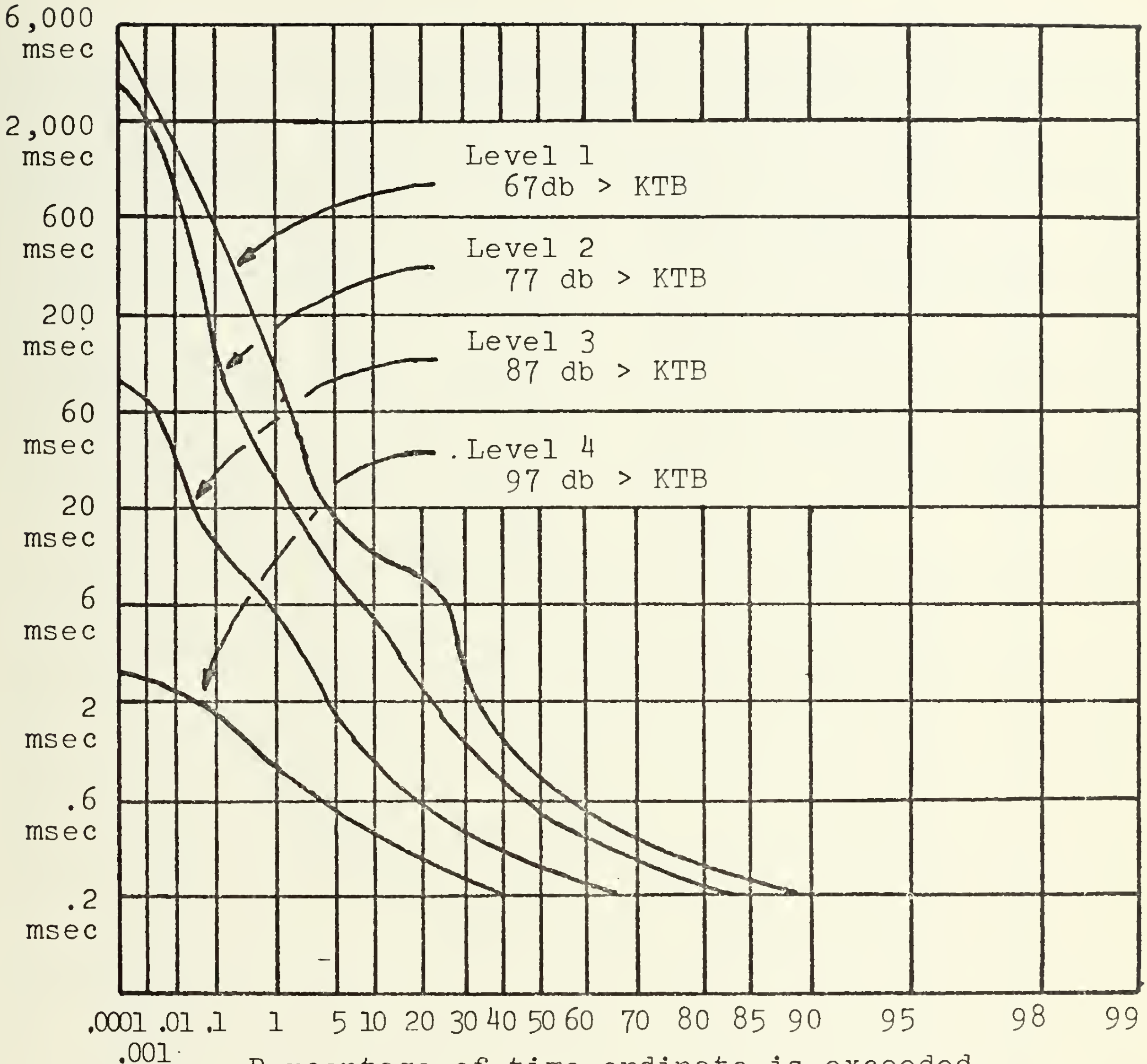

Figure 2.7. Pulse Duration Distribution of Measured atmospheric noise. Recorded on 4 April 1969 between 0000 and 0400 MST at Boulder, Colorado at frequency 5 MHz.

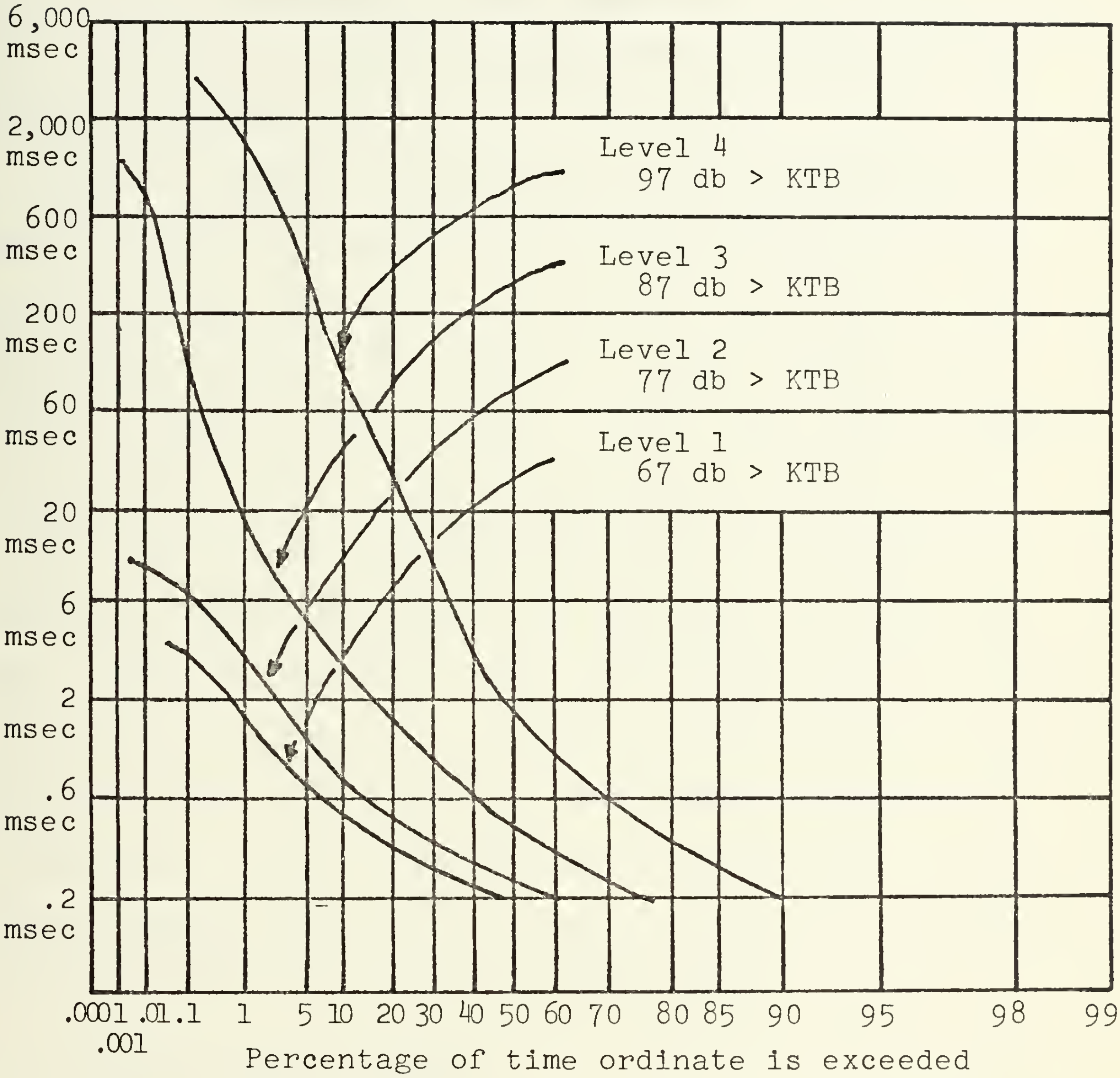

Figure 2.8. Pulse interval distribution of atmospheric noise. Recorded on 4 April, 1969 between 0000 and 0400 MST at Boulder, Colorado on a frequency of 5 MHz.

plots of the pulse duration distribution and pulse interval distribution respectively, for atmospheric noise at 5.0 MHz. For completeness, the APD for this set of samples is also plotted as Figure 2.9.

Since the median signal energy to noise energy ratio is nominally greater than 20 db for the HF channel, we will use this figure to obtain estimates of the effect of the atmospheric noise on the channel. It may be seen from Figure 2.9 that a 20 db signal to noise ratio will correspond roughly to a bit error probability approximately 5×10^{-5} while Figure 2.7 indicates that for the pulse duration distribution only 7% of the pulses were greater than .6 msec in duration. (The 20 db signal to noise ratio corresponds to level 5.) A similar analysis of Figure 2.8 indicates that 55% of the intervals between bursts were longer than 60 msec. For many HF circuits, particularaly where a large signal to noise ratio exists, the atmospheric noise introduces essentially random errors.

D. MODELING THE HF CHANNEL

The preceding sections developed a relation between the bit error probabilities and the number of multipaths in existance between the transmitter and receiver. The behavior observed is not new or unexpected; it has long been known that as few as six equal amplitude signals of random phase may be summed to give a distribution which is nearly Rayleigh. We must, however, bear in mind the assumptions utilized to arrive at the results obtained: recall that we assumed

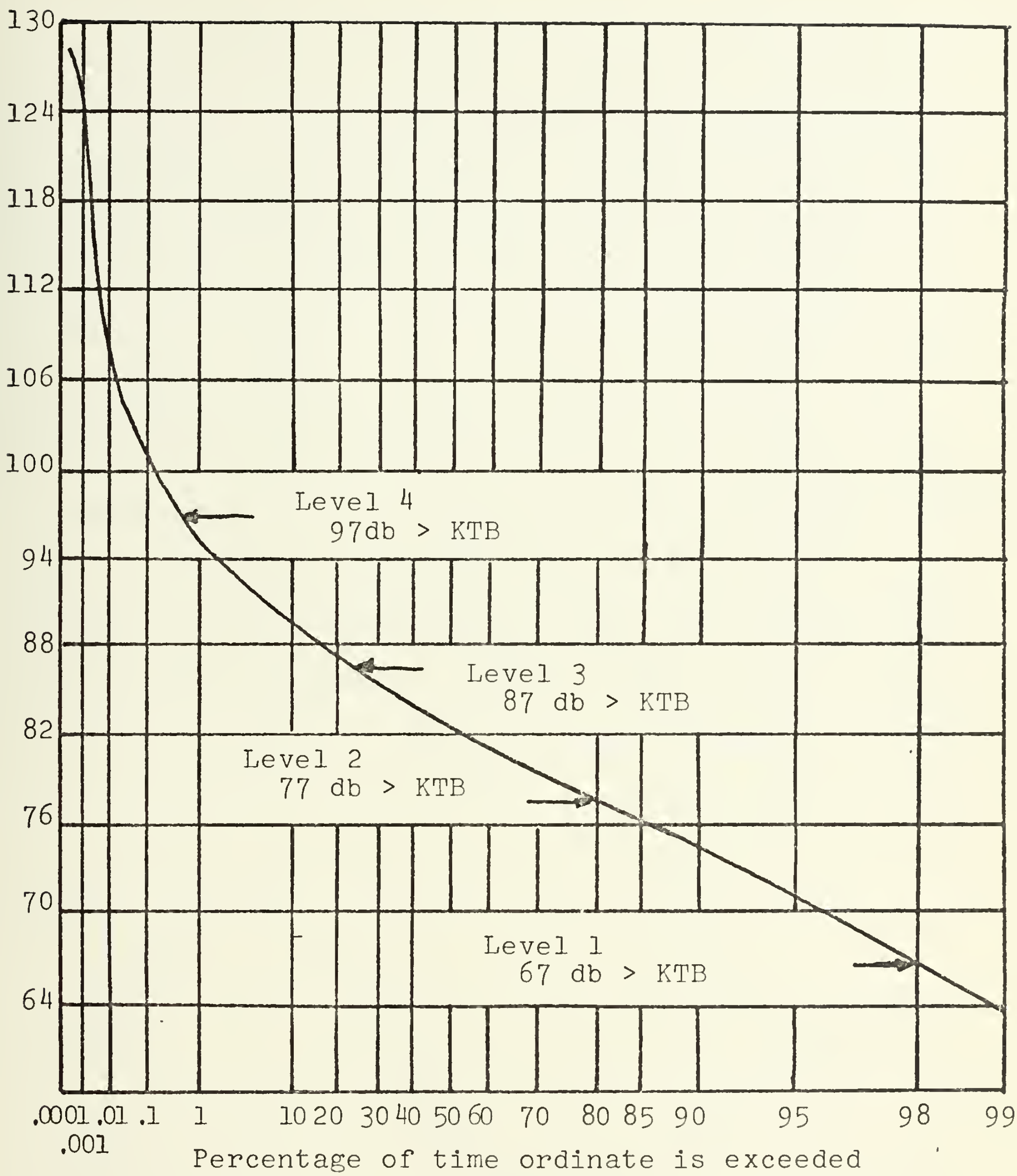

Figure 2.9. APD of atmospheric noise measured at Boulder, Colorado 0000- 0400 MST, 4 April 1969, at a frequency of 5.0 MHz.

stationary and statistically independent specular and scatter components. It should be mentioned that some of the multipath conditions may result from magneto-ionic splitting and may be correlated. In addition we did not discuss the time delay (which may be as much as several milliseconds) nor the frequency distortion or doppler. In general we are dealing with a channel where there is correlation between errors, and even groups of errors. In graphical terms the bit error probability may be expressed as a function of time as in Fig. 2.10. Time periods, when the error probability is relatively high, could well result from the existance of multiple propagation paths, while the periods of small error probability are most likely due to little or no multipath and only atmospheric noise corrupting the signal.

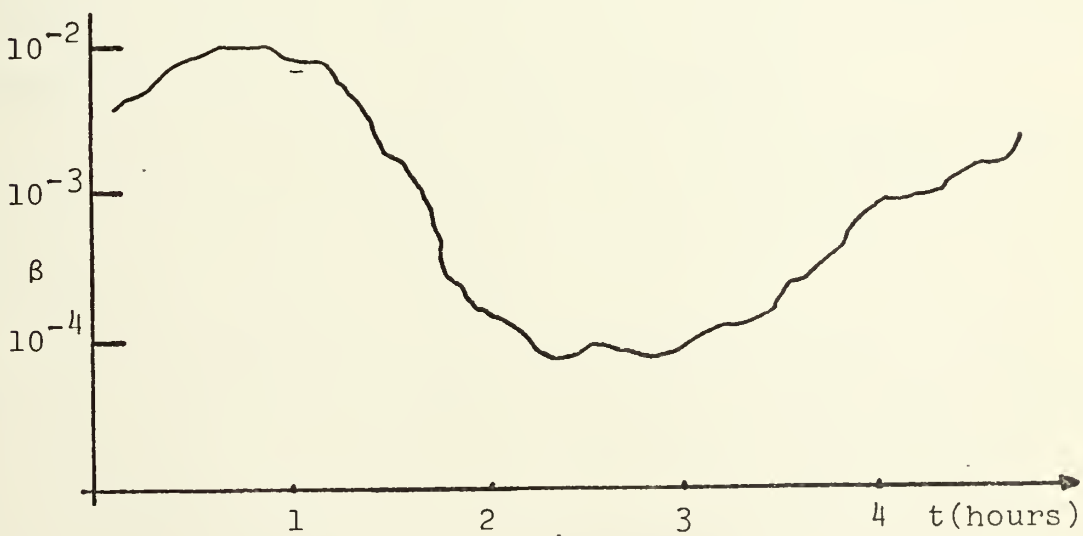

Figure 2.10. Error Probability as a Function of Time.

Another factor which limits the use of the channel is the differential time delay incurred by the signal traveling different paths. This time delay places a lower limit on the signal duration (for pulsed systems) since, if the signal duration is about equal to the time delay over one of the propagation paths, it will be received during the subsequent signal time frame. This differential time delay is on the order of 5 msec and hence a lower limit on the signal duration is between 5 and 10 msec. To make more efficient use of the allocated channel bandwidth, the normal procedure is to frequency multiplex several channels onto a common carrier and utilize a signal duration of about 10 msec on each of the sub-channels. Again propagation anomilities such as the doppler shift may cause 'cross talk' between adjacent channels. This, in many instances, results in frequency selective fading where the fading is confined to a small section of the transmitted bandwidth.

In order to describe the channel in qualitative terms we need to know how the errors occur. The simple bit error probability is not sufficient.

If the message is to be transmitted as binary digits suitably modulating a radio frequency carrier, then a burst of length N is defined as a sequence of N digits beginning with a bit in error that is immediately preceded by a correct bit, and ending with an error which is followed by a correct bit. The burst interval is defined as the interval between bursts. It begins with the first correct digit following a

burst and ends with the last correct digit preceding the next burst. A gap of length m is defined to be a sequence of m consecutive error free digits between two successive errors. A gap may occur within a burst or actually be the burst interval.

It is apparent that the sequence of received digits may be completely punctuated by the errors into bursts and burst intervals, however this partitioning is not unique. The burst length and burst interval distributions are statistics of the channel and must either be known or at least a reasonable statistical estimate should be available. Since the partitioning is not unique, the burst and burst interval should not be used to characterize the channel. To avoid this difficulty, we return to the gap distribution. While the gap distribution will affect the burst length, it is not subject to interpretation ambiguities and may be used to uniquely determine a possible channel model.

There are two primary methods of characterizing the HF channel. The first consists of actually recording error patterns and statistics for a reasonably extensive period of time and assuming these are representative of the channel. The data may then be utilized to determine the effectiveness of various coding schemes by actual simulation or the performance estimated by the coding parameters. The second approach is to use, again, measured parameters to infer a stationary homogenous Markov chain model which may then be utilized to characterize the channel. The usefulness of

this method follows from the properties of the stationary, homogeneous Markov chain, since the chain is completely characterized by the matrix of one step probability transitions. This approach is presented in detail by Fritchman [13], Tsai [5], and McManamon [6], together with results which are in very good agreement with measured data. The aperiodic Markov chain model, as described by McManamon, is completely determined by the gap distribution. Since a gap of length n is, by definition, a sequence of n consecutive error free digits which are immediately preceded and followed by errors, the probability of a gap of length n, f(n) may be interpreted as the return to an error state in exactly n+1 steps. Therefore, f(n) is simply the first return probability distribution of the Markov chain. The mean return time may then be expressed as

$$\mu = \sum_{n=1}^{\infty} nf(n)$$

The stationary distribution of the Markov chain is of considerable practical importance. Denoting the probability that the output of the channel at time j is in error by P(j), we may then ask 'How does P(j) behave as time progresses?' The answer is given by the Ergodic Theorem for Markov chains (Cox and Miller [14]), where the stationary distribution is defined in terms of the mean return time. The standard definition for the first return in the Markov chain includes the return state. Therefore, in terms of the gap distribution f(n), as defined above,

$$\lim_{j \to \infty} P(j) = \frac{1}{\mu + 1} = \beta$$

where β is the average bit error probability. This simply states that after a long period of time the average probability of the output being in error is independent of the initial distribution and is equal to the inverse of the average gap distribution.

In summary, the HF channel is neither the classical random error channel nor the typical burst channel, but rather a hybrid channel which may contain may contain many long bursts between which occur random errors. For the convenience of the reader, the measured burst and burst interval distributions given by Tsai [5] are reproduced here as Figures 2.11 and 2.12 respectively.

Figure 2.11 may be used to obtain the burst duration, for example 80 percent of the bursts are less than 20 bits in duration. Figure 2.12 gives the interval between bursts, and by a similar examination it can be seen that 40 percent of the burst intervals were less than 1200 bits.

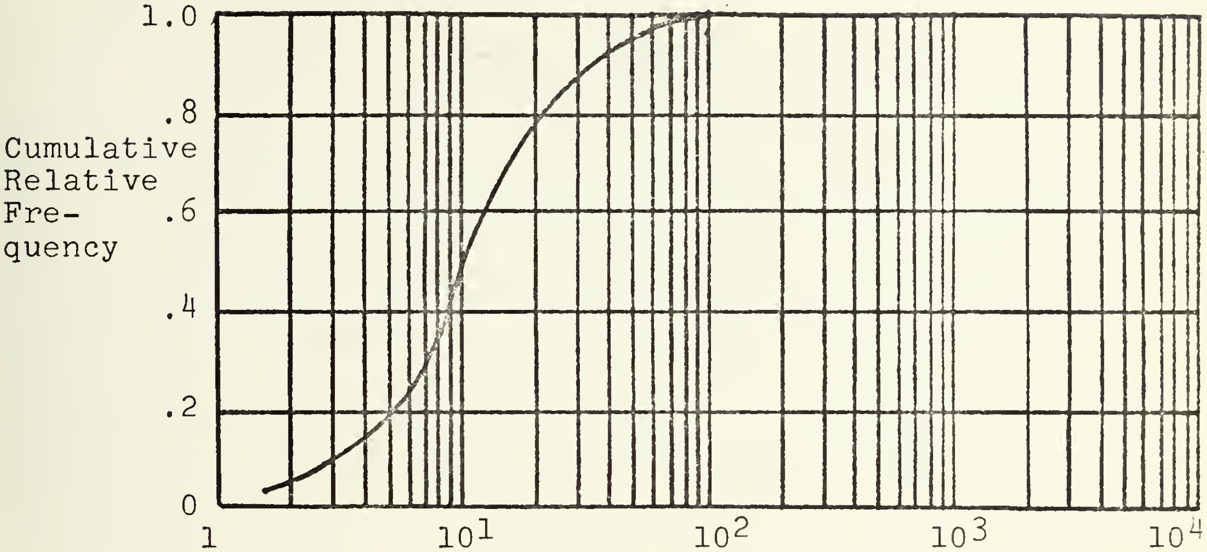

Figure 2.11. "Burst distribution for an HF channel"

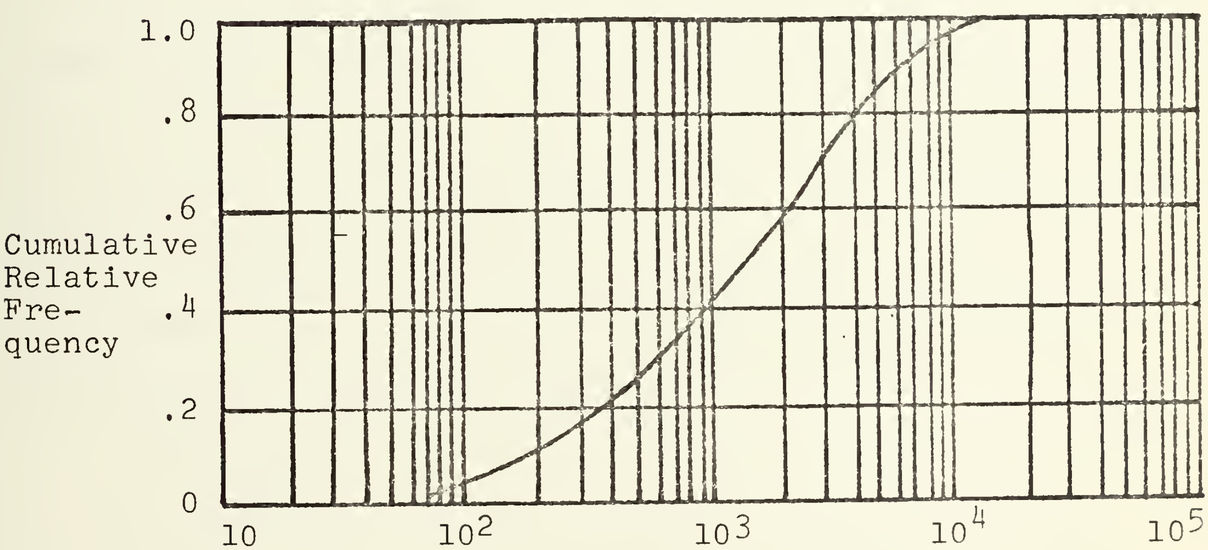

Figure 2.12. "Burst interval distribution for HF channel.

III. CODING

A. INTRODUCTION

In the preceding chapter the effect of the channel on the transmitted information was described. It should be noted that the modulation/detection procedure assumed is non-coherent frequency-shift-keying. This assumption was made since it is in wide use and of considerable practical importance.

Before we discuss specific coding procedures let us first review, in a very basic manner, what we want to do, and what is available. A discrete, or digital communication system in which the performance may be described by error rates and probabilities is assumed. The first question that comes to mind is "Is there some 'source' which we could utilize in the channel to obtain the maximum rate, R, of communication with the smallest error rate?". The answer is yes; for a significant class of channels, Shannon has shown that a non-zero capacity, C, exists such that

$$C = \max_{[\text{source}]} R$$

with an arbitrarily low error rate. In general it is not possible to engineer a system to the point where R = C, and the probability of an error is extremely small. We choose to modify the question and ask "For some rate R (< C), can we significantly reduce the probability of an error?". In other words we are not really trying to solve the

"optimum" problem but rather a problem of what is reasonable.

If we assume for the moment that we have a return channel, we have available simple error detection techniques, such as ARQ (Automatic Repeat Request) and the parity check schemes. The ARQ systems, introduced by Van Durren, are frequently referred to as 3-out-of-7, or "constant ratio codes," where the encoded messages consist of seven bits in which three 'ones' and four 'zeros' are required to form a valid code word. Any single error, and errors of odd multiplicity (3,5, and 7), will be detected while only half of the even errors will be detected. The remaining even errors will then constitute 'real' errors at the output.

In the parity check system a sequence of n-1 information carrying digits are appended by a single digit, known as the parity check digit, so that the total number of ones in the n sequence is even. Again at the receiver, the parity is checked and if this check fails, an error has been detected. It is again apparent that this scheme will detect all errors of odd multiplicity. The efficiency is much higher being $\frac{n-1}{n}$, however, the rate again suffers by the retransmission.

Another alternative, which is more attractive, is a system where error correction is possible. In general, the error correcting code which will correct 't' errors, consists of a mapping of k information digits into n channel digits in such a fashion that at least 2t + 1 channel errors

will be required to change any valid transmitted word into another valid word. The decoding may be done in many ways. One, becoming more promising as the state of the art develops, is by a table look-up from a read-only-memory. In this procedure the received word is compared with the valid code words and the one which differs in the fewest positions is then chosen as the most likely one to have been sent.

In the above paragraphs we tacitly assumed that any set of n channel digits were determined by a unique set of k information digits, and conversely one set of k information digits uniquely determined one set of n channel digits; this corresponds to block coding. Another form of coding permits any set of k information digits to affect several blocks of channel digits, and conversely any block of channel digits will be affected by several blocks of information digits; this corresponds to convolutional (or recurrent) coding.

If the i^{th} sequence of n channel digits contains the corresponding i^{th} sequence of information bits explicitly, the code is said to be systematic. Systematic codes may have a marked advantage over non-systematic codes in that they may, of course, be decoded with error correction by those with the decoding equipment, while those with only a limited amount of equipment may, with proper synchronization, simply regenerate the code from the appropriate information digits and use the result for error detection alone, relying

on a retransmission to correct the indicated errors. Those who have little equipment may, for the price of timing, have access to the information digits alone.

In summarizing the above procedures, we must again look at the channel. If the operational requirements are such that a two way, or return circuit is not possible (for example in a fleet broadcast network) the error correction capability is highly desired.

B. CONVOLUTIONAL CODES

Let us begin by describing a few items which are very basic to the discussion which follows. The term encoding is used to imply a transformation or mapping from the set of information sequences into another set of channel sequences, as illustrated in Figure 3.1. The code is the image of the transformation.

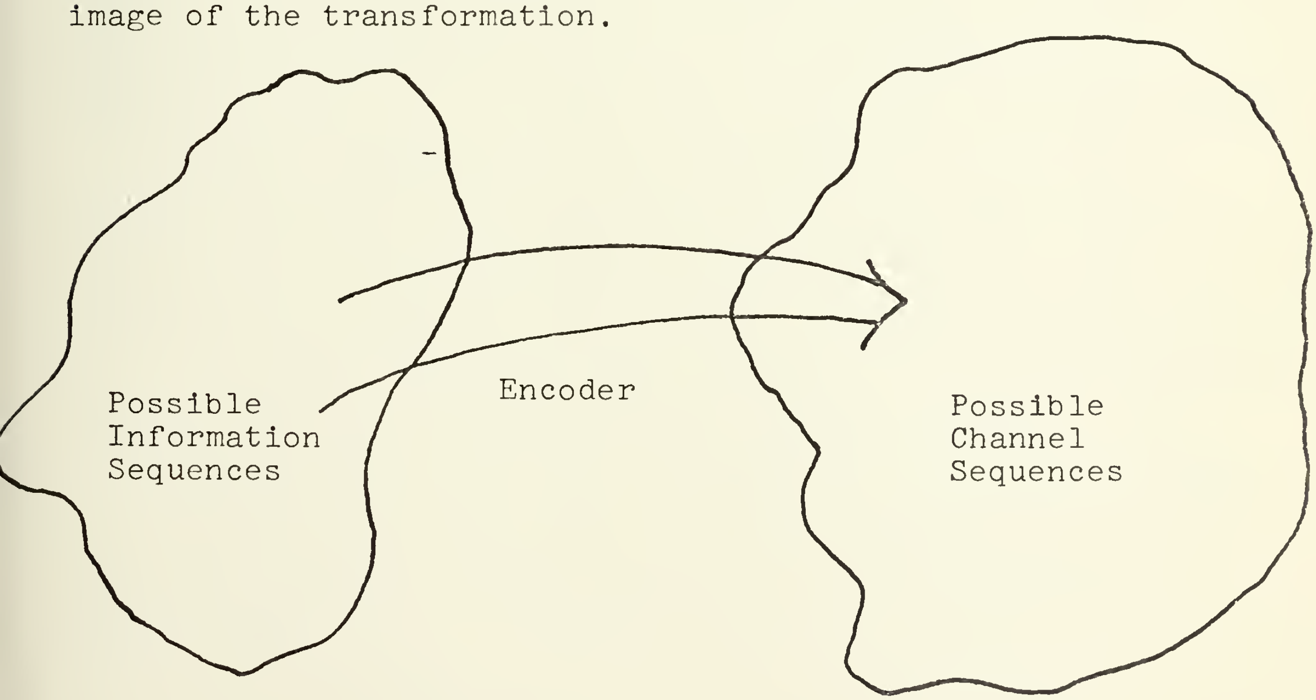

Figure 3.1. Mapping provided by code.

We restrict our attention to binary codes, where the elements of the information and channel sequences are either 0, or 1. If each input sub-block contains α digits, while the corresponding channel sub-block contains γ digits, the code is said to be a rate $\frac{\alpha}{\gamma}$ code and designated as a "(γ,α) code." Since we are dealing with binary codes, the addition we will use is Modulo 2, where

$$0 + 0 = 1 + 1 = 0$$

and

$$1 + 0 = 0 + 1 = 1$$

A convolutional code is said to have <u>memory span</u> ν, if each encoded digit is a linear combination of information digits in the present set, and at most $\nu - 1$ of the preceding sets of information digits. It should be noted in passing, that if $\nu = 1$, the code corresponds to a block code.

Another parameter, which is closely related to the memory span, is the <u>actual constraint length</u> (N), which is defined for a convolutional code of memory span ν, as $N = \nu\gamma$. This constraint length plays a role in convolutional codes similar to the block length in block codes.

For Convolutional codes, we define $t_{i,j}$ as the j^{th} digit of the i^{th} set of channel digits, in terms of message digits $m_k(1)$, as

$$t_{i,j} = \sum_{q=0}^{\nu-1} \sum_{n=1}^{\alpha} g_{q,j}(n) m_{i-q}(n) \qquad 1 \leq j \leq \gamma \quad \text{(Modulo 2)} \tag{22}$$

where $m_{i-q}(n)$ is the n^{th} digit of the $(i-q)^{th}$ set of information digits, and $g_{q,j}(n)$ is either 0, or 1 depending on whether or not $m_{i-q}(n)$ is to be included in the j^{th} transmitted digit. From equation 22, we can see that any particular message digit can affect at most ν sets of γ transmitted digits, which prompts the definition of memory span.

Turning our attention to the elements $g_{q,j}(n)$, the choice of values for these elements will clearly determine the properties of the code. If we define $g_{q,j}(n)$ as:

$$g_{0,j}(n) = \ _{j,n} \qquad 1 \leq j \leq \alpha$$

$$g_{q,j}(n) = 0 \qquad 1 \leq j \leq \alpha \quad q \neq 0$$

where $\delta_{j,k}$ is zero if $j \neq k$, and 1 if $j = k$. The first α digits of the i^{th} set of channel digits will then be identical with the α digits of the i^{th} set of information digits, in which case the code is said to be <u>systematic</u> (or canonic systematic). In most of the following we will treat systematic codes only, since Massey [15] has shown that any non-systematic code can be converted to an equivalent systematic code without affecting the error correcting capability of the code. In general the constraint length of the non-systematic code will be shorter than that for the equivalent systematic code.

At this point a short example might help to further explain the procedures.

<u>Example 1</u>. A Systematic (3,1) convolutional code with memory span 5. From equation 22 we have

$$t_{i,j} = \sum_{q=0}^{\nu-1} \sum_{n=1}^{\alpha} g_{q,j}(n) m_{i-q}(n). \qquad \text{(E1-1)}$$

Since the desired code has $\alpha = 1$, we let $m_{i-q}(1) = m_{i-q}$, and $g_{q,j}(n) = g_{q,j}$, we may rewrite equation E1-1 as

$$t_{i,j} = \sum_{q=0}^{\nu-1} g_{q,j} m_{i-q}.$$

Now, since the memory span $\nu = 5$, we have $\nu - 1 = 4$ and:

$$t_{i,j} = g_{0,j}m_i + g_{1,j}m_{i-1} + g_{2,j}m_{1-2} + g_{3,j}m_{i-3} + g_{4,j}m_{i-4}.$$

Since the code is systematic, we know

$$g_{0,1} = 1; \quad \text{and} \quad g_{q,1} = 0.$$

Let us choose, somewhat arbitrarily,

$$g_{0,2} = g_{1,2} = g_{4,2} = 1; \quad g_{2,2} = g_{3,2} = 0 \qquad \text{(E1-2)}$$

and

$$g_{0,3} = g_{2,3} = 1; \quad g_{1,3} = g_{3,3} = g_{4,3} = 0.$$

The code now may be written as

$$t_{i,1} = m_i$$

$$t_{i,2} = m_i + m_{i-1} + m_{i-4}$$

$$t_{i,3} = m_i + m_{i-2}$$

For this code the actual constraint length is $N = \gamma\nu = 3x5 = 15$.

A different method of visualizing this encoding procedure was developed by Wyner and Ash [16], and consists in considering $g_{q,j}(1),\ldots,g_{q,j}(\alpha)$ to be a row vector as

$$G_{q,j} = [g_{q,j}(1)\ g_{q,j}(2)\ \cdots\ g_{q,j}(\alpha)]$$

and M_{i-q} a column vector of the $(i-q)^{th}$ set of information digits such that

$$M_{i-q} = [m_{i-q}(1)\ m_{i-q}(2)\ \cdots\ m_{i-q}(\alpha)]^T.$$

Then

$$t_{i,j} = \sum_{q=0}^{\nu-1} G_{q,j} M_{i-q}.$$

If we now let T_i be the column vector of the i^{th} set of transmitted digits, we have

$$T_i = \begin{bmatrix} t_{i,1} \\ t_{i,2} \\ \cdot \\ \cdot \\ \cdot \\ t_{i,\gamma} \end{bmatrix} \begin{bmatrix} G_{\nu-1,1} & \cdots & G_{1,1} & G_{0,1} \\ G_{\nu-1,2} & \cdots & G_{1,2} & G_{0,2} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ G_{\nu-1,\gamma} & \cdots & G_{1,\gamma} & G_{0,\gamma} \end{bmatrix} \begin{bmatrix} M_{i-\nu+1} \\ M_{i-\nu+2} \\ \cdot \\ \cdot \\ \cdot \\ M_i \end{bmatrix}.$$

By dealing only with systematic codes we may greatly simplify the matrix, since $g_{0,j}(n) = \delta_{j,n}$ and $g_{q,j}(n) = 0$ for $q \neq 0$ for j between 1 and α.

$$\left[\begin{array}{cc|ccc|c} 0 & 0 & \cdots & & 0 & I \\ \hline G_{\nu-1} & G_{\nu-2} & \cdots & & G_1 & G_0 \end{array}\right] \begin{bmatrix} M_{i-\nu+1} \\ M_{i-\nu+2} \\ \cdot \\ \cdot \\ \cdot \\ M_i \end{bmatrix}$$

Where I is the $(\alpha x \alpha)$ identity matrix, 0 is the (α,α) null matrix and G_j is the $(\gamma-\alpha)$ x α matrix defined as

$$
G_j = \begin{bmatrix} G_{j,\alpha+1} \\ G_{j,\alpha+2} \\ \vdots \\ G_{j,\gamma} \end{bmatrix} = \begin{bmatrix} g_{j,\alpha+1}{}^{(1)} & g_{j,\alpha+1}{}^{(2)} & \cdots & g_{j,\alpha+1}{}^{(\alpha)} \\ g_{j,\alpha+2}{}^{(1)} & g_{j,\alpha+2}{}^{(2)} & \cdots & g_{j,\alpha+2}{}^{(\alpha)} \\ \vdots & & & \\ g_{j,\gamma}{}^{(1)} & g_{j,\gamma}{}^{(2)} & \cdots & g_{j,\gamma}{}^{(\alpha)} \end{bmatrix} \tag{24}
$$

If the number of input message sequences is indefinitely long, we may consider the transmitted channel digits to form a semi-infinite column vector T. Where T

$$
T = \begin{bmatrix} T_1 \\ \\ T_2 \\ \\ \vdots \\ T_{\nu-1} \\ \\ T_\nu \\ \\ T_{\nu+1} \\ \\ T_{\nu+2} \\ \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} I & 0 & 0 & & & & & & & & \\ G_0 & \phi & \phi & & & & & & & & \\ 0 & I & 0 & & & & & & & & \\ G_1 & G_0 & \phi & & & & & & & & \\ \vdots & \vdots & & & & & & & & & \\ 0 & 0 & 0 & \cdots & 0 & I & 0 & 0 & 0 & 0 & 0 \\ G_{\nu-2} & G_{\nu-3} & G_{\nu-4} & & G_1 & G_0 & \phi & \phi & \phi & \phi & \phi \\ 0 & 0 & 0 & & 0 & 0 & I & 0 & 0 & 0 & 0 \\ G_{\nu-1} & G_{\nu-2} & G_{\nu-3} & & G_2 & G_1 & G_0 & \phi & \phi & \phi & \phi \\ 0 & 0 & 0 & & 0 & 0 & 0 & I & 0 & 0 & 0 \\ \phi & G_{\nu-1} & G_{\nu-2} & & G_3 & G_2 & G_1 & G_0 & \phi & \phi & \phi \\ 0 & 0 & 0 & & 0 & 0 & 0 & 0 & I & 0 & 0 \\ \phi & \phi & G_{\nu-1} & & G_4 & G_3 & G_2 & G_1 & G_0 & \phi & \phi \\ & & & \ddots & & & & & & & \end{bmatrix}}_{G_\infty} \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \\ \vdots \\ \\ \\ M_{-1} \\ \\ M_{-} \\ M_{+1} \\ \\ M_{+2} \\ \vdots \end{bmatrix} \tag{25.}
$$

and ϕ is a $(\gamma - \alpha) \times \alpha$ null matrix. For convenience we will refer to the semi-infinite matrix which multiplies the column vector of message digits as G_∞, while the matrix

which generates only the i^{th} set of channel digits may be defined from equation 23 as

$$G = \left[\begin{array}{c|c|c|c|c} 0 & 0 & \dots & 0 & I \\ \hline G_{\nu-1} & G_{\nu-2} & \dots & G_1 & G_o \end{array}\right] \tag{26}$$

Although G_∞ has an infinite number of rows and columns, any particular row will have at most $\alpha\nu$ non-zero elements and, correspondingly, any particular column will have at most $N = \gamma\nu$ non-zero elements.

Returning to example 1, from equation 23 we may write the i^{th} transmitted sequence as

$$T_i = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{i-4} \\ m_{i-3} \\ m_{1-2} \\ m_{i-1} \\ m_i \end{bmatrix}$$

and by inspection, equations 25 and 26 yield

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$G_\infty = \begin{bmatrix} 100000000 \ldots \\ 100000000 \ldots \\ 100000000 \ldots \\ 010000000 \ldots \\ 110000000 \ldots \\ 010000000 \ldots \\ 001000000 \ldots \\ 011000000 \ldots \\ 011000000 \ldots \\ 101000000 \ldots \\ 000100000 \ldots \\ 001100000 \ldots \\ 010100000 \ldots \\ 000010000 \ldots \\ 100110000 \ldots \\ 001010000 \ldots \\ 000001000 \ldots \\ 010011000 \ldots \\ 000101000 \ldots \\ 000000100 \ldots \\ 001001100 \ldots \\ 000010100 \ldots \\ \ddots \end{bmatrix}$$

The physical generation of this code is straight forward involving unit delay elements (shift registers) and modulo 2 adders (exclusive 'OR' gates), as shown in Figure 3.2. The rectangular boxes represent unit delay elements, while the circles with the cross represent the modulo 2 adders.

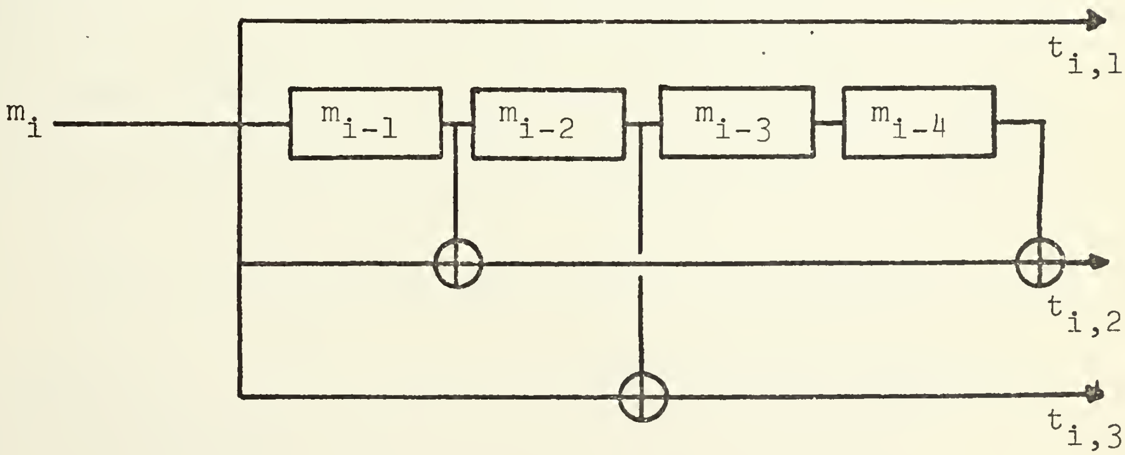

Figure 3.2. Encoder for Rate $\frac{1}{3}$ convolutional code

Returning to the general problem, when the channel digits are transmitted through the channel, noise may corrupt some of them, causing the received digits to differ from the original transmitted digits. Utilizing matrix notation, the i^{th} received sequence may be expressed as:

$$R_i = \begin{bmatrix} r_{i,1} \\ r_{i,2} \\ \cdot \\ \cdot \\ \cdot \\ r_{i,} \end{bmatrix} = \begin{bmatrix} t_{i,1} \\ t_{i,2} \\ \cdot \\ \cdot \\ \cdot \\ t_{i,} \end{bmatrix} + \begin{bmatrix} e_{i,1} \\ e_{i,2} \\ \cdot \\ \cdot \\ \cdot \\ e_{i,} \end{bmatrix} = T_i + E_i \qquad (27)$$

Where $e_{i,j}$ is 1 if an error has occurred in the j^{th} position of the i^{th} transmitted sequence and zero otherwise. Our task may simply be stated as determining which element of E_i are most likely to be non-zero.

Since we are dealing with systematic codes, we have available estimates of the i^{th} set of information digits, namely $m'_{i,j} = r_{i,j}$. By utilizing the i^{th} $\nu - 1$ preceding sets, we may regenerate the code with a duplicate encoder and obtain an estimate of the transmitted parity digits as:

$$t'_{i,j} = \sum_{q=0}^{\nu-1} \sum_{k=1}^{\alpha} g_{q,j}(k) r_{i-q,k} \qquad \alpha+1 \leq j \leq \gamma$$

By adding $t'_{i,j}$ to $r_{i,j}$ ($\alpha+1 \leq j \leq \gamma$) we obtain what is called the <u>syndrome</u> $s_{i,j}$.

$$s_{i,j} = t'_{i,j} + r_{i,j}$$

$$s_{i,j} = \sum_{q=0}^{\nu-1} \sum_{k=1}^{\alpha} g_{q,j}(k)(m_{i-q,k} + e_{i-q,k})$$

$$+ \sum_{q=0}^{\nu-1} \sum_{k=1}^{\alpha} g_{q,j}(k)m_{i-q,k} + e_{i,j}$$

since

$$t_{i,j} = \sum_{q=0}^{\nu-1} \sum_{k=1}^{\alpha} g_{q,j}(k)m_{i-q,k} .$$

Since the encoder is linear we may further reduce this syndrome to

$$s_{i,j} = \sum_{q=0}^{\nu-1} \sum_{k=1}^{\alpha} g_{q,j}(k)(m_{i-q,k} + m_{i-q,k})$$

$$+ \sum_{q=0}^{\nu-1} \sum_{k=1}^{\alpha} g_{q,j}(k)\, e_{i-q,k} + e_{i,j}$$

however, $m_{i-q,k} + m_{i-q,k} = 0 \quad \text{Mod } 2.$

Therefore

$$s_{i,j} = \sum_{q=0}^{\nu-1} \sum_{k=1}^{\alpha} g_{q,j}(k)e_{i-q,k} + e_{i,j} \quad \text{for } \alpha + 1 \leq j \leq \gamma \tag{28}$$

For each set of γ received digits we will be able to form $\gamma - \alpha$ syndromes, which depend only on the errors in the received sequence.

Again let us consider the matrix notation and form a column matrix of syndromes. From equation 28 we may form S_i as

$$\hat{S}_i = \begin{bmatrix} S_{i,\alpha+1} \\ S_{i,\alpha+2} \\ S_{i,\alpha+3} \\ \cdot \\ \cdot \\ \cdot \\ S_{i,\gamma} \end{bmatrix} = \begin{bmatrix} G_{\nu-1} & 0 & G_{\nu-2} & 0 & \cdots & 0 & G_1 & 0 & G_0 & \underline{I} \end{bmatrix} \begin{bmatrix} E_{i-\nu+1} \\ E_{i-\nu+2} \\ \cdot \\ \cdot \\ \cdot \\ E_i \end{bmatrix} \tag{29}$$

Where 0 is a$(\gamma-\alpha)x(\gamma-\alpha)$ null matrix, G_j is defined in equation 24, and $\underline{I}$ is an $(\gamma-\alpha)x(\gamma-\alpha)$ identity matrix. We are now able to write the entire set of syndromes, S, in a fashion similar to equation 25.

$$S = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \cdot \\ \cdot \\ \cdot \\ S_{\nu-1} \\ S_\nu \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \underbrace{\begin{bmatrix} \overbrace{G_0 \quad \underline{I}}^{\gamma \text{ columns}} & \phi & 0 & \phi & \cdot & \cdot & \cdot & & & & & & & \\ G_1 \quad 0 & G_0 & \underline{I} & \phi & 0 & \cdot & \cdot & \cdot & & & & & & \\ G_2 \quad 0 & G_1 & 0 & G_0 & \underline{I} & \phi & 0 & \cdot & \cdot & \cdot & & & & \\ & & \cdot & & & & & & & & & & & \\ & & & \cdot & & & & & & & & & & \\ & & & \cdot & & & & & & & & & & \\ G_{\nu-1} \quad 0 & G_{\nu-1} & \cdot & \cdot & \cdot & 0 & G_1 & 0 & G & \underline{I} & \phi & 0 \\ \phi \quad 0 & \bar{G}_{\nu-1} & \cdot & \cdot & \cdot & 0 & G_2 & 0 & G_1 & 0 & G_0 & \underline{I} \\ & & & & \cdot & & & & & & & \\ & & & & \cdot & & & & & & & \\ & & & & & \cdot & & & & & & \end{bmatrix}}_{A_\infty} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ \\ \\ \\ E_{\nu-1} \\ E_\nu \\ \\ \\ \\ \end{bmatrix}$$

or

$$S = A_\infty E \; .$$

C. THRESHOLD DECODING

With these syndromes available, we turn our attention to a form of decoding originally introduced by Massey [15] called Threshold Decoding, in which the error decoding decision is based on the value of the majority of the

syndromes corresponding to those which would be '1' if $e_{i,j}$ were actually in error. From equation 30 we see the first set of syndromes affected by E_i (i.e. the i^{th} set of error digits) is S_i. Since the memory span is ν, we know the effects of errors in E_i will have to be confined to the syndromes S_i - $S_{i+\nu-1}$, hence:

$$\begin{bmatrix} S_i \\ S_{i+1} \\ S_{i+2} \\ \vdots \\ S_{i+\nu-1} \end{bmatrix} = \begin{bmatrix} G_0 & 1 \\ G_1 & 0 \\ G_2 & 0 \\ \vdots & \vdots \\ G_{\nu-1} & 0 \end{bmatrix} \begin{bmatrix} E_i \end{bmatrix}$$

Where as before G_k is a $(\gamma-\alpha) x \alpha$ matrix whose elements are either 0 or 1, i.e. the j^{th} column of G_k is

$$\begin{matrix} g_{k,\alpha+1}(j) \\ g_{k,\alpha+2}(j) \\ g_{k,\alpha+3}(j) \\ \vdots \\ g_{k,\gamma}(j) \end{matrix}$$

Thus those syndromes, which will be affected by the error in the j^{th} position of the error sequence E_i will be those syndromes for which $g_{l,k}(j) = 1$, $(\alpha+1 \leq k \leq \gamma;\ j \leq l \leq j+\nu-1)$. In other words, if $e_{i,j} = 1$ and no other errors occur, only those syndromes which have $g_{l,k}(j) = 1$ will be non-zero. The presence of other errors within the sets of $\nu - 1$ error sequence either prior to, or after E_i may alter some of

these syndromes to negate the effect of the error. The decoding proceeds by examining those syndromes which corresponds to $e_{i,j}$ and a 'majority vote' decides on the value of $e_{i,j}$, i.e., if more than half of the syndromes are 1 the decision is that $e_{i,j} = 1$ and the corresponding information digit is changed. If the digit is corrected the effect of the error may be removed from the set of syndromes by inverting all of those syndromes used in the calculation.

This mode of operation is referred to as feedback decoding, and as long as no decoding error is made, the decoder operation is superior to decoding without feedback (normally called definite decoding). On the other hand, when a decoding error is made, the eroneous correction tends to propagate. This error propagation is one of the acknowledged difficulties of convolutional codes. To avoid error propagation we introduce the concept of a self-orthogonal (or canonic self-orthogonal) code due to Robinson and Bernstein [17]. A self-orthogonal code is a convolutional code in which any two syndromes which check $e_{i,j}$ have at most one error digit in common. Since all the syndromes which check $e_{i,j}$ have $e_{i,j}$ in common, the definition requires any other error digit to be included in at most one syndrome. Robinson [18] has shown that a self-orthogonal code has limited error propagation, and on the other hand, may be decoded without feedback with no error propagation with only a slight loss in error correction capability.

To illustrate the threshold decoding procedures, recall example 1, where

$$T_i = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{i-4} \\ m_{i-3} \\ m_{i-2} \\ m_{i-1} \\ m_i \end{bmatrix}$$

By comparison with equation 26 we have

$$G_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix};\ G_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix};\ G_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix};\ G_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix};\ G_4 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

As in equation 11 the syndromes which check the i^{th} digit may be seen from:

$$\begin{bmatrix} s_{i,2} \\ s_{i,3} \\ s_{i+1,2} \\ s_{i+1,3} \\ s_{i+2,2} \\ s_{i+2,3} \\ s_{i+3,2} \\ s_{i+3,3} \\ s_{i+4,2} \\ s_{i+4,3} \end{bmatrix} = - \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} e_{i,1} \\ e_{i,2} \\ e_{i,3} \end{bmatrix}$$

and we see $e_{i,1}$ will effect the syndromes $s_{i,2}$; $s_{i,3}$; $s_{i+1,2}$; $s_{i+2,3}$; and $s_{i+4,2}$. With five syndromes, the majority rule decides $e_{i,1} = 1$ if and only if three or more of these syndromes are 1. If this set of five syndromes is expressed in terms of the error digits, it can be seen that the only digit in common between any two is $e_{i,1}$, hence

this code is self-orthogonal. The decoder is illustrated in Figure 3.3 where as before, the rectangular boxes are delay elements and the circles are modulo 2 adders.

Returning to the general concept of threshold decoding, it should be pointed out that the encoder and decoder are relatively simple to implement. To obtain the maximum benefit from threshold decoding, the code should be self-orthogonal. Robinson and Bernstein [17] have developed a synthesis procedure, utilizing finite difference sets whereby these self-orthogonal codes may be constructed.

Confining our attention to systematic codes of rate one half, the transmitted digits assume the form

$$t_{i,1} = m_{i,1}$$

$$t_{i,2} = \sum_{k=0}^{\nu-1} g_k m_{i-k}$$

or, in matrix notation

$$T_i = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ g_{\nu-1} & g_{\nu-2} & \cdots & g_1 & g_0 \end{bmatrix} \begin{bmatrix} m_{i-\nu+1} \\ \vdots \\ m_{i-2} \\ m_{i-1} \\ m_i \end{bmatrix}$$

After the digits have been transmitted and received, we may again form the syndromes, realizing for this case we will have only one syndrome for each set of transmitted digits. Forming the i^{th} syndrome by analogy with equation 29

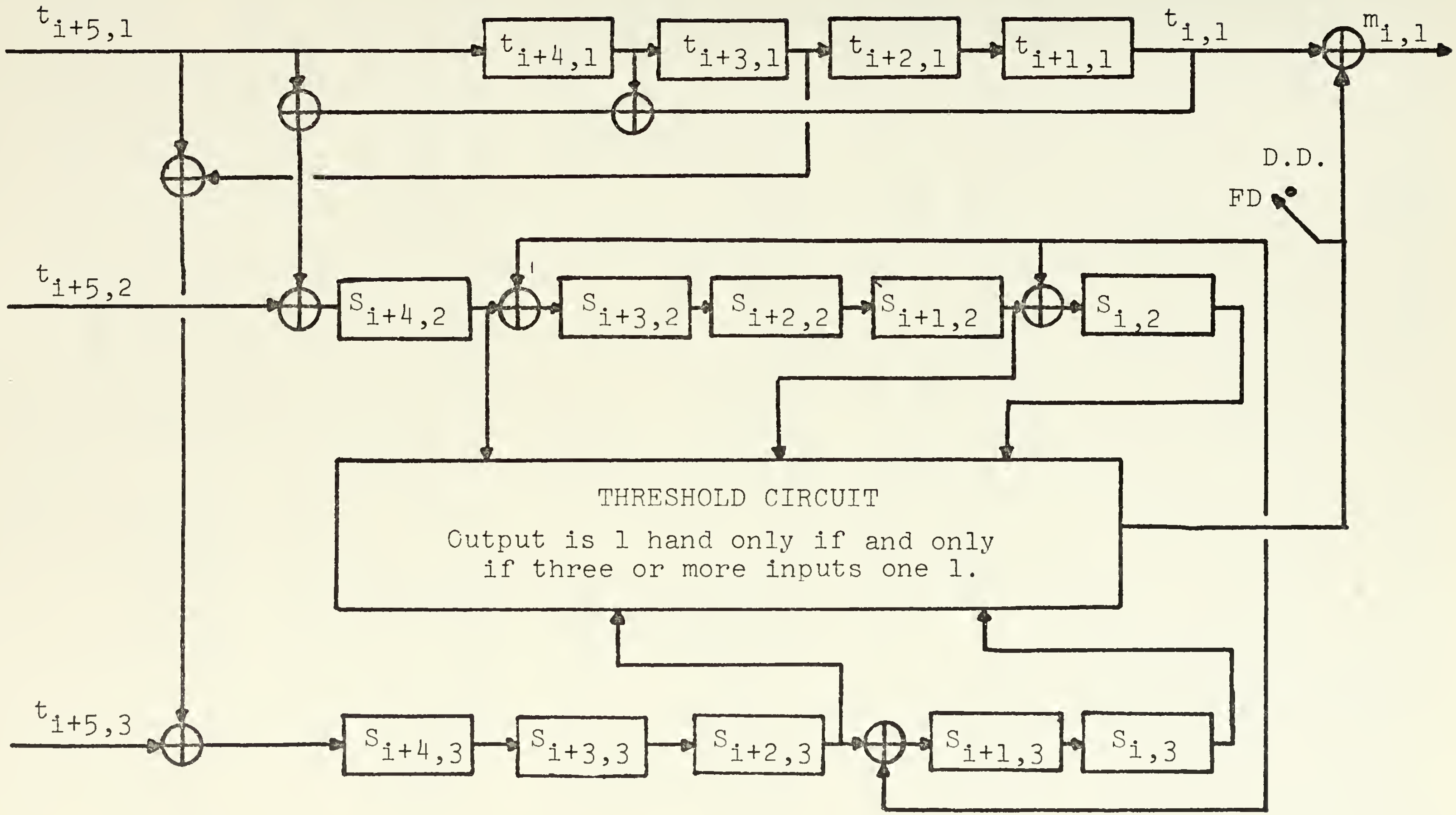

Figure 3.3. Decoder for rate $\frac{1}{3}$ convolutional code

$$s_i = \begin{bmatrix} g_{\nu-1} & 0 & g_{\nu-2} & 0 & \cdots & g_1 & 0 & g_0 & 1 \end{bmatrix} \begin{bmatrix} e_{i-\nu+1,1} \\ e_{i-\nu+1,2} \\ \vdots \\ e_{i-1,1} \\ e_{i-1,2} \\ e_{i,1} \\ e_{i,2} \end{bmatrix}$$

We note that only the information error digits are included in the i^{th} syndrome calculation except for the parity error digit $e_{i,2}$, hence for convenience we will isolate the information and parity error digits and write instead

$$S_i = \begin{bmatrix} s_i \\ s_{i+1} \\ \cdot \\ \cdot \\ \cdot \\ s_{i+\nu-1} \end{bmatrix} = \overbrace{\begin{bmatrix} g_{\nu-1} & g_{\nu-2} & \cdots & g_1 & g_0 & 0 & 0 & 0 \\ 0 & g_{\nu-1} & \cdots & g_2 & g_1 & g_0 & 0 & 0 \\ & & \ddots & & & & & \\ 0 & 0 & \cdots & & g_{\nu-1} & g_{\nu-2} & \cdots & g_0 \end{bmatrix}}^{A} \begin{bmatrix} e_{i-\nu+1,1} \\ e_{i-\nu+2,1} \\ \cdot \\ \cdot \\ \cdot \\ e_{i+\nu-1,1} \end{bmatrix} + \begin{bmatrix} e_{i-\nu+1,2} \\ e_{i-\nu+2,2} \\ \cdot \\ \cdot \\ \cdot \\ e_{i+\nu-1,2} \end{bmatrix}$$

Where S_i is a column matrix of the syndromes $s_i, \ldots, s_{i+\nu-1}$. From the definition of memory span, the i^{th} message digit can only affect these syndromes. Examining the ν^{th} column of the matrix, it is obvious that $e_{i,1}$ will affect only those syndromes for which $g_j = 1$.

For threshold decoding we examine only those syndromes for which the ν^{th} column of A contains $g_j = 1$, and form a reduced set of syndromes

$$S_{Ri} = \begin{bmatrix} g_{\nu-1} & g_{\nu-2} & \cdots & g_1 & g_0 & 0 & 0 & 0 & & \\ & & & & \vdots & & & & & \\ & & & & g_{\nu-1} & \cdots & & g_0 & & \end{bmatrix} \begin{bmatrix} e_{i-\nu+1,1} \\ \vdots \\ e_{i+\nu-1,1} \end{bmatrix} + \begin{bmatrix} e_{i-\nu+1,2} \\ \vdots \\ e_{i+\nu-1,2} \end{bmatrix}$$

By stipulating the code to be self-orthogonal, we are assured that no column (other than the ν^{th}) will contain more than one element.

If there are a total of L syndromes which check $e_{i,1}$, then the majority logic decides $e_{i,1} = 1$, if and only if the number of non-zero syndromes is greater than $[\frac{L}{2}] + 1$, where $[\frac{L}{2}]$ is the integer part of $\frac{L}{2}$. A relatively straight forward calculation shows there are $\frac{L}{2}(L-1)$ previously decoded error positions, $\frac{L}{2}(L-1)$ error positions which have not yet been decoded, L parity error positions, and $e_{i,1}$, which will effect the decoding, for a total of $L^2 + 1$ possible positions for errors to occur. If definite decoding (i.e. no feedback correction to the syndromes) is used, we may calculate the probability of making a decoding error for the binary symmetric channel (where the probability of making an error is β independent of the original value of the transmitted digit).

The probability of correctly decoding $e_{i,1}$ is simply the probability that $[\frac{L}{2}]$ or fewer errors occur in the possible $(L^2 - 1)$ positions, hence

$$P_{(\text{correct decoding})} = \sum_{i=0}^{[\frac{L}{2}]} \binom{L^2+1}{i} \beta^i (1-\beta)^{L^2+1-i}$$

Now

$$\sum_{i=0}^{[\frac{L}{2}]} \binom{L^2+1}{i} \beta^i (1-\beta)^{L^2+1-i} = 1 - \sum_{i=[\frac{L}{2}]+1}^{L^2+1} \binom{L^2+1}{i} \beta^i (1-\beta)^{L^2+1-i} ,$$

and the probability that a bit is decoded in error is P_{DD}

$$P_{DD} = 1 - P_{(\text{correct decoding})}$$

$$= \sum_{i=[\frac{L}{2}]+1}^{L^2+1} \binom{L^2+1}{i} \beta^i (1-\beta)^{L^2+1-i} .$$

If β is reasonably small this may be approximated by the first term as

$$P_{DD} \simeq \binom{L^2+1}{[\frac{L}{2}]+1} \beta^{[\frac{L}{2}]+1} (1-\beta)^{L^2+1-([\frac{L}{2}]+1)} . \tag{33}$$

The decoding error probability for feedback decoding, on the other hand, is more difficult to analyze. The bit under consideration will be decoded correctly if i previously decoded bits were in error, provided that fewer than $[\frac{L}{2}] - i$ errors occur in the bits which have not yet been decoded and the L syndrome error positions. Hence the probability that a bit is decoded correctly is

$$P_{(\text{correct decoding})} = \sum_{i=0}^{[\frac{L}{2}]} P_{(\text{undecoded errors} \leq \frac{L}{2} - i \,|\, i \text{ decoding errors})} \cdot P_{(i \text{ decoding errors})} \tag{34}$$

For the conditional probability there are $\frac{L}{2}(L+1)$ total undecoded error positions, hence, assuming independent errors,

$$P_{(\text{undecoded errors} \leq \frac{L}{2} - i \mid i \text{ decoding errors})}$$

$$= \sum_{j=0}^{[\frac{L}{2}]-i} \binom{\frac{L(L+1)}{2}}{j} \beta^j (1-\beta)^{\frac{L(L+1)}{2} - j} \quad . \qquad (35)$$

Designating the probability of a decoding error as P_{FD} (for feedback decoding) and assuming again independent errors within the $\frac{L}{2}(L-1)$ previously decoded positions,

$$P_{(i \text{ decoding errors})} = \binom{\frac{L(L-1)}{2}}{i} P_{FD}^{i} (1-P_{FD})^{\frac{L}{2}(L-1)-1} \qquad (36)$$

Therefore,

$$P_{(\text{correct decoding})} = \sum_{i=0}^{[\frac{L}{2}]} \sum_{j=0}^{[\frac{L}{2}]-1} \binom{\frac{L(L+1)}{2}}{j} \beta^j (1-\beta)^{\frac{L(L-1)}{2} - j}$$

$$\binom{\frac{L(L-1)}{2}}{i} P_{FD}^{i} (1-P_{FD})^{\frac{L(L-1)}{2} - i}$$

Noting that $1 - P_{(\text{correct decoding})} = P_{FD}$, this may be reduced to

$$1 = \sum_{i=0}^{[\frac{L}{2}]} \left\{ \sum_{j=0}^{[\frac{L}{2}]-1} \binom{\frac{L(L+1)}{2}}{j} \beta^j (1-\beta)^{\frac{L(L+1)}{2} - j} \right\} \binom{\frac{L(L-1)}{2}}{i} P_{FD}^{i}$$

$$(1 - P_{FD})^{\frac{L(L-1)}{2} - i-1} \qquad (37)$$

Although rather tedious, by hand, the solution may be obtained readily on a computer. Obvious approximations may be made, when β is sufficiently small so that $P_{DD} << 1$, and $P_{FD} < P_{DD}$ a very reasonable approximation is obtained by assuming $P_{FD}^3 \approx 0$, and solving the quadratic equation which results.

D. OTHER DECODING METHODS

1. Sequential Decoding

Another algorithm for decoding convolutional codes is sequential decoding, which was introduced by Wozencraft [19], and differs from threshold decoding in that it is probabilistic in nature. The encoder for a (γ,α) code is viewed as a directed graph or tree composed of nodes and branches. For a binary (γ,α) code, from each node 2^{α} branches emanate, and to each branch is assigned a sequence of γ binary digits. The decoder has available or is able to calculate the channel probability characteristics, and a copy of the tree which the encoder is utilizing. When the first few sequences are received the decoder proceeds to hypothesize which sequences were 'most probably' sent and moves one node farther into the tree. This process continues until errors occur, at which time the decoder may get off onto the wrong branch. When this occurs it is forced to make additional errors, since the sequences on the incorrect path have no relation to the sequences actually transmitted. The decoder will eventually realize

that something is wrong, at which time it back-tracks to a previously visited node and tries another branch. If none of these branches are better, it again has to back-track to yet another previous node and search the branches eminating from that node. All of these operations take time, in addition to the time required to calculate the channel probability characteristics, when this must be done. If there is only a finite amount of buffer storage available for the incoming data, a long search might not be completed before the storage registers are filled, in which case a buffer overflow occurs and the system must be resynchronized.

Sequential decoding may be done on any convolutional code and, if the constraint length is very, very long, the decoder will eventually decode the transmitted sequences correctly with probability approaching 1.

For illustration purposes consider a simple (3,1) systematic convolutional code of memory span 3. Figure 3.4 is the tree graph of the desired code.

At each node are two branches corresponding to the two possible values of the information digit (i.e. 0 or 1), a '0' implies the upper branch while a '1' implies the lower branch. This code was constructed by utilizing the procedures developed by Lin and Lyne [20].

The first input sub-block (in this case a single digit) determines which of two output sub-blocks will be transmitted, i.e., 1 → 111; 0 → 000. The second output sequence will depend on both the first and second input sequences.

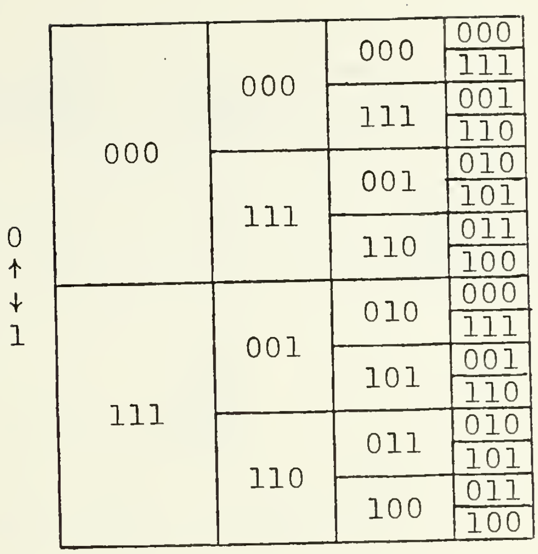

Figure 3.4. Tree graph at convolutional code.

If the first input was '1' and the second '0', then the second output sequence would be 001. The third output sequence, in a like fashion, will depend on the first two inputs and the third input. The fourth output sequence is dependent on the fourth input and the preceding two inputs - and so the process continues. As an example the imput sequence 10110 would be mapped into 111 001 100 110 011. The encoder for this code may be implemented with unit delay elements and modulo 2 adders as indicated in Figure 3.5.

In the decoding procedure, the receiver examines three sets of sequences, then outputs the best estimate of the first digit. For example, assume the transmitted sequence were 111 001 101 111 corresponding to an input 1 0 1 1, while the received sequences were 100 001 101 110 011.

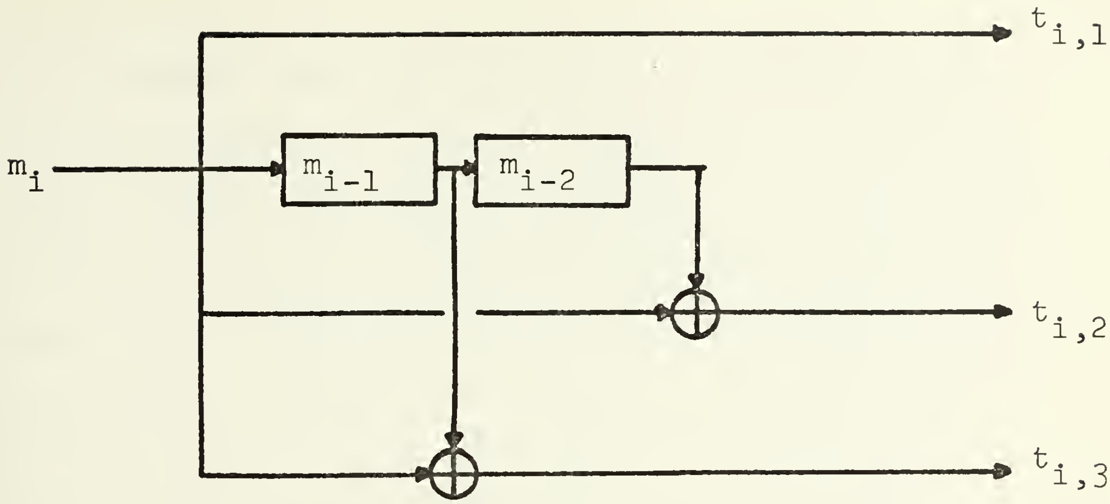

Figure 3.5. Encoder for convolutional tree code shown in Figure 3.4

The decoder 'knows' that the first sequence was either 111 or 000, between these two 000 is the more likely (assuming a memoryless binary symmetric channel where one error is more likely than two).

Therefore, the initial hypothesis is that the first digit of the transmitted sequence is '0', the decoder then proceeds to the next node, again the two choices are 000, and 111, of these 000 is again the most logical since only one error is required to transform 000 into 001. Going on to the third node the decoder chooses 111 again with one error. At this point the total number of errors required to transform the transmitted sequence into the received sequence is three. The decoder calculates at this point (as at the other nodes) the set of channel probability characteristics and decides that perhaps three errors in the nine digits is excessive in which case it looks at the other branches from the third node. The other

branch is quickly ruled out since four errors would be indicated. The decoder then goes back to the second node and looks at the other branch, where again the number of errors would be three, so the decoder continues back to the first node. Again the first node lower branch indicates two errors, however, the decoder goes along this path to the second node, where the upper branch has no indicated errors. Going farther into the tree to the third node there are again no additional errors hence, the decoder alters its initial hypothesis and assumes the first bit to be '1'. Proceding into the fourth node there are no indicated errors within the constraint length, so the decoder hypothesis is the second digit is '0'. And so the process continues, backtracking when the number of of errors becomes excessive.

2. Viterbi Decoding Scheme

Another method of decoding convolutional codes is the Viterbi algorithm [21]. Omura [22] has shown that this decoding scheme may be viewed as a forward dynamic programming solution to a generalized regulator problem. The state of the encoder at time k+1 is dependent on the state at time k and the k^{th} message input, hence one may write for the state at time k+1

$$X_{k+1} = A\ X_k + B\ M_k$$

The output of the encoder at time k is some function of the state of the encoder at time k, hence the output may be

denoted as $T_k = G_k(X_k)$. Again transmitting the encoder output through the channel, the received sequence will be $R_k = T_k + E_k$. Defining a performance measure J(T) as

$$J(T) = 1 \log P(R|T) = - \sum_{k=0}^{L} \log P(R_k|T_k)$$

the decoding proceeds by defining intermediate costs as

$$f_{L-k}(X_k) = \min_{\substack{(M_j) \\ j=k}}^{L-1} - \left\{ \sum_{j=k+1}^{L} \log P(R_j|T_j) \right\}$$

for each k = 0,1,...,L-1, and defining $f_0(X_L) = 0$. Since the decoder follows the path with the minimum cost an error will occur if and only if the minimum cost is incurred on some path other than the correct path.

E. APPLICATION OF CODING TO CHANNEL

The maximum expected differential time delay of 5 msec for different propagation paths requires the signal to have a duration of about 10 msec, which in turn implies a maximum data rate of about 100 bits per second. For convenience, we shall assume that the data rate to be used is 75 bits per second, which corresponds to the standard 100 word per minute teletype. In addition, we shall use a rate ½ code, and, since the encoder outputs two digits for each input digit, the modulation method will be frequency division multiplexing the information digits onto one 75 bit per second subchannel and the parity digits onto another. A block diagram for the system is given in Figure 3.6.

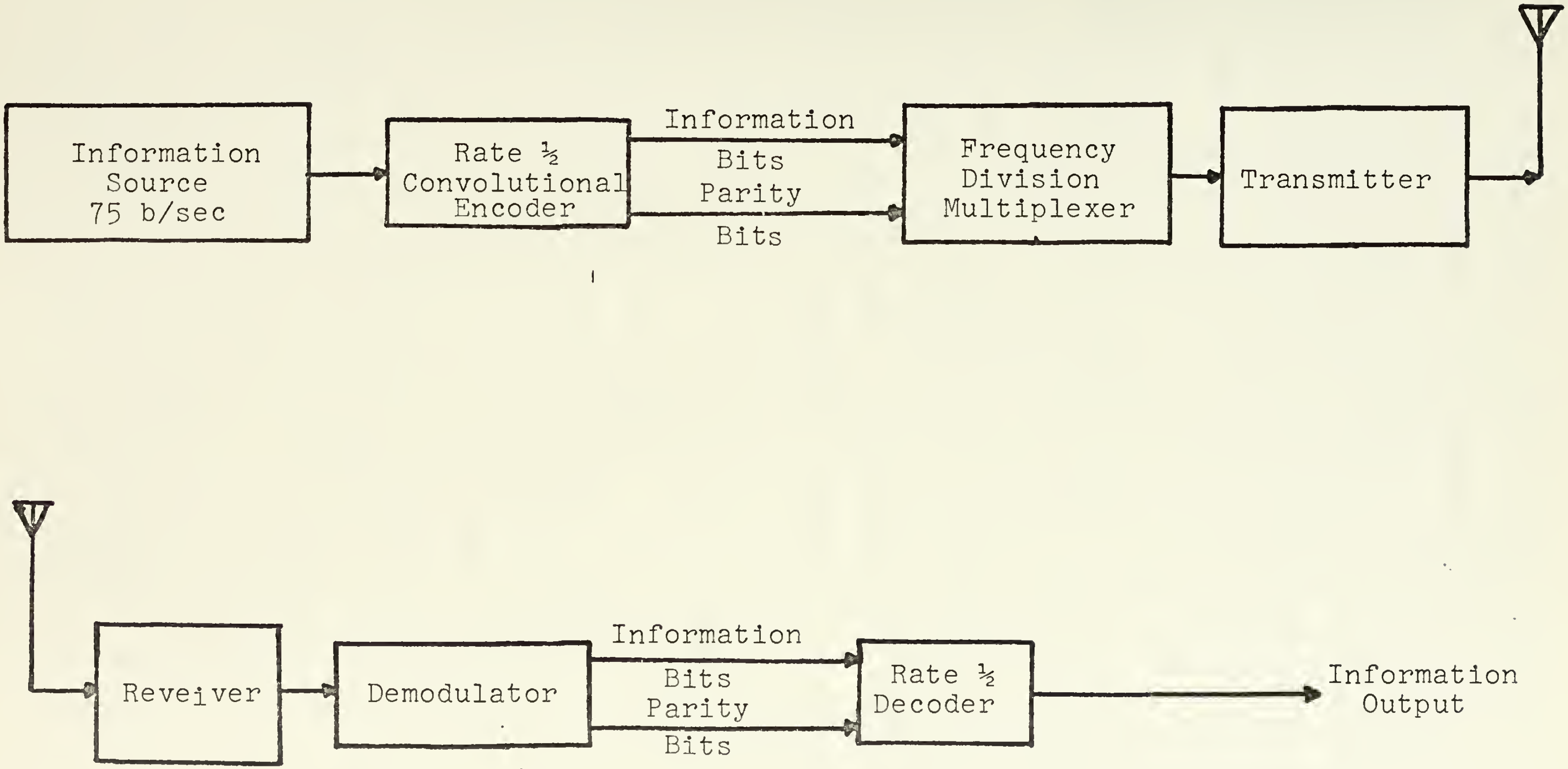

Figure 3.6. Block diagram for implementing a rate ½ code

The distributions given by Tsai [5] were obtained utilizing a channel data rate of 75 bits per second, and indicate that with probability 0.97, the bursts are less than 50 bits in duration. While it is not impossible to implement a burst error correcting code to decode bursts of length 50, the burst correcting codes all require a minimum guard space of error free digits. This error free guard space is not generally found in the HF channel.

Since burst correcting codes are not desirable, the alternative is to modify the distribution of the bursts. This burst modification may be performed by a process known as interleaving. Interleaving is simply a time ordering of the transmitted digits so that adjacent errors in the channel are spread out in the received sequence prior to decoding. (Some forms of interleaving may be viewed as time multiplexing.) Interleaving by k bits may be accomplished by many methods, however, one of the simplest is to replace each delay element in both the encoder and decoder by a k bit shift register. This has the effect of inserting k columns of zeros between the columns in the generator matrix, G_∞, of the code, and correspondingly k rows of zeros between each of the rows in the syndrome matrix, A_∞. Any burst of errors that is k digits or shorter in length will not affect more than one of the digits in each of k succeeding decoding steps. On the other hand, there may be errors from more than one

burst included in the decoding calculation. The interleaving cannot alter the average bit error probability, all it does is rearrange the errors so they occur in a more evenly distributed fashion. For the data presented by Tsai, interleaving by 50 bits would then insure that fewer than two error digits from a particular burst would be contained in the digits utilized for any decoding operation with probability 0.97.

The amount of interleaving required is dependent on the duration and frequency of the bursts, and the constraint length and correction capabilities of the code being used. As the amount of interleaving is increased the channel will appear more and more random, however, it must be noted that the interleaved data will be extremely susceptible to periodic interference unless care is used in choosing the interleaving. (For example 64 bit interleaving would not be desirable for use with a modulator utilizing 32 subchannels, since occasional modulator synchronization errors would be related to the interleaving. This would have a much more serious effect on the proper decoding than if the interleaving were chosen to be 63 or 65 bits.)

When attempting to ascertain the degree of interleaving required, actual channel characteristics or measured error patterns are utilized. For a given set of received digits, varying amounts of interleaving may be tried to determine when the errors are "random enough." To determine what "random enough" implies, consider a set of data for which

the average bit error probability is β. If the constraint length of the code is N and the code is interleaved by I bits, then the effective constraint length will be N·I, and the decoding will be performed by examining this effective constraint length. Although the total number of errors is still the same, the ratio of the number of errors per block approaches β, since, by the law of large numbers, the probability of having more than β NI errors in NI digits approaches zero as NI increases indefinitely. In general this would imply an excessively large amount of interleaving, since the trend may be rather slow.

Another criterion is available to determine a reasonable estimate for the amount of interleaving. For any given code, the ratio of the number of errors correctable to the number of possible error positions may be obtained. For the definite decoding procedure of the preceding section this is $\frac{[\frac{L}{2}]}{L^2+1}$. The decoder will function properly (or at least with some improvement) until the ratio of the number of errors to the number of possible error positions reaches $\frac{[\frac{L}{2}]}{L^2+1}$. Hence to determine a practical interleaving degree, choose I so that the ratio of the number of blocks of length NI with more than $\frac{[\frac{L}{2}]}{L^2+1}$ NI errors, to the total number of blocks is approximately equal to the probability of a decoding error for a completely random channel with the error probability β. For convenience we will use the definite decoding error probability P_{DD}. The term

approximately above is used loosely since the effectiveness of the interleaving is extremely sensitive to the actual bursts received.

To illustrate the procedure, consider a particular code capable of correcting two errors, with constraint length 18 as given by:

$$T_i = \begin{bmatrix} t_{i,1} \\ t_{i,2} \end{bmatrix} = \begin{bmatrix} 000000001 \\ 100010110 \end{bmatrix} \begin{bmatrix} m_{i-8} \\ m_{i-7} \\ \cdot \\ \cdot \\ \cdot \\ m_i \end{bmatrix}$$

Interleaving by I bits will simply insert I columns of zeros between each of the columns of the generator matrix above, and replace m_{j-k} by m_{j-kI}. The reduced set of syndromes on which the threshold decoding is based are:

$$S_{Ri} = \begin{bmatrix} s_{(i+I)} \\ s_{(i+2I)} \\ s_{(i+4I)} \\ s_{(i+8I)} \end{bmatrix} = \begin{bmatrix} 100010110000000 \\ 010001011000000 \\ 000100010110000 \\ 000000010001011 \end{bmatrix} \begin{bmatrix} e_{(i-8I),1} \\ \cdot \\ \cdot \\ \cdot \\ e_{i,1} \\ \cdot \\ \cdot \\ \cdot \\ e_{(i+8I),1} \end{bmatrix} + \begin{bmatrix} e_{(i+1I),2} \\ e_{(i+2I),2} \\ e_{(i+4I),2} \\ e_{(i+8I),2} \end{bmatrix}$$

Again I rows of zeros are to be inserted between each of the rows of the multiplying matrix above. For this code there are a total of 17 possible error positions within which 2 errors may be corrected hence the ratio of correctable errors

to total error positions is $\frac{2}{17}$. From equation 33, the decoding error probability, P_{DD}, may be approximated as

$$P_{DD}(\beta) = \binom{17}{3} \beta^3(1-\beta)^{14}$$

$$\simeq 680\beta^3 .$$

We now turn our attention to the actual error data, as contained in reference 23. We will consider three typical cases of measured error statistics where the average error rate β is about $2.5\text{x}10^{-2}$. This implies that P_{DD} is about $5\text{x}10^{-3}$. Interleaving by I = 27 bits results in the number of blocks with more than $\frac{2}{17}$ (18)(27) equal to 95 out of a total of 63675 blocks for a 'residual' uncorrectable error rate of $1.5\text{x}10^{-3}$. Interleaving by 54 bits lowers this to about $1.2\text{x}10^{-3}$, while interleaving by 104 bits only brings it down to $7.86\text{x}10^{-4}$. Continuing to increase the interleaving to the point where I = 855 only changes the residual error rate to $5.9\text{x}10^{-4}$. This is presented in tabulation form relative to NI in Table III. Again it should be mentioned that the effectiveness of the interleaving is strongly dependent on the actual characteristics of the bursts. (This information is not available for the data presented.) It appears as though a conservative amount of interleaving would be 54 bits for the code under consideration, since for this value, the code would correct many of the channel errors, and still be reasonably practical.

Although the code presented here is not a very powerful one, as compared with some of the block codes or longer convolutional codes, the ratio of correctable errors to total possible errors is nearly as good as the Golay block code - which is capable of correcting 3 errors - ($\frac{2}{17}$ as compared with $\frac{3}{24}$). In addition, the simple syndrome decoding procedure is flexible enough so that by monitoring the syndrome register, the mode of decoding (definite, or feedback) could be determined - choosing feedback decoding during periods when there are relatively few errors indicated in the syndrome register and switching to definite decoding when the number of indicated errors could imply possible error propagation. This would give a higher output error probability during definite decoding, however there would be no error propagation.

For the feedback decoding, and reasonably small β, P_{FD} is at least as small as P_{DD}, and we assume $P_{FD}^3 \approx 0$.

From equation 35

$$P(\text{channel errors} \leq 2 | 0 \text{ decoding errors}) \approx 1 - \binom{11}{3}\beta^3 = 1 - 165\beta^3$$

$$P(\text{channel errors} \leq 1 | 1 \text{ decoding error})\ 1 - \binom{11}{2}\beta^2(1-\beta)^9 = 1 - 55\beta^2 + 99\beta^3$$

$$P(\text{channel errors} = 0 | 2 \text{ decoding error}) (1-\beta)^{11} = 1 - 1\beta + 55\beta^2 - 165\beta^3$$

and from equation 36:

$$P(0 \text{ decoding errors}) = (1-P_{FD})^6 = 1 - 6P_{FD} + 15P_{FD}^2$$

$$P(1 \text{ decoding error}) = 6P_{FD}(1-P_{FD})^5 = 6P_{FD} - 30P_{FD}^2$$

$$P(2 \text{ decoding errors}) = 15P_{FD}^2$$

Substituting these expressions into equation 38 and solving the resulting quadratic equation yields

$$P_{FD} = \frac{1+330\beta^2-1584\beta^3}{30(-11\beta+165\beta^2-518\beta^3} \left(1 \pm \sqrt{1 - 2 \frac{165\beta^3-(11\beta+165\beta^2+518\beta^3)30}{(1+330\beta^2-1584\beta^3)^2}}\right)$$

By expanding the square root and retaining only the first two terms (since the third term involves $((\beta^4)^2)$), P_{FD} becomes

$$P_{FD} = \left(\frac{1+330\beta^2-1584\beta^3}{30(-11\beta+165\beta^2-518\beta^3)}\right)\left(1 \pm \left(1 - \frac{165\beta^3 x 30(-11\beta+165\beta^2+518\beta^3)}{(1+330\beta^2-1584\beta^3)^2}\right)\right)$$

The minus sign is to be used, since the positive sign results in a probability greater than 1 for small values of β; and P_{FD} may be simplified to

$$P_{FD} \approx \frac{165\beta^3}{1+330\beta^2}$$

A comparison of the decoding error probabilities is given in Table II.

TABLE II

β	P_{DD}	P_{FD}
10^{-2}	$6.80 x 10^{-4}$	$1.65 x 10^{-4}$
10^{-3}	$6.80 x 10^{-7}$	$1.65 x 10^{-7}$
10^{-4}	$6.80 x 10^{-10}$	$1.65 x 10^{-10}$

It must be noted that these calculations are for a completely random channel, however by interleaving by 54 or

more bits they should be within an order of magnitude of the observed statistics, again depending to a large extent on the actual distribution of errors.

The discussion of effectiveness of coding must include some mention of block coding performance. One of the better block codes is the modified Golay [24, 12] code, which will correct 3 errors within any 24 bit code word. For reasonably small β the probability of a decoding error P_{GC} is approximately given as

$$P_{GC} \simeq \binom{24}{4}\beta^4$$

$$\simeq 10626\beta^4$$

The error probability P_{GC}, is <u>not</u> a bit error probability, but the probability for a 24 bit word, or equivalently a 12 bit information word. To be comparable with the bit error probability obtained for the convolutional code, this word error probability must be converted to a bit error probability [Ref. 23]. For the degrees of interleaving utilized for the convolutional code these bit error probabilities are listed in Table III.

As can be seen from Table III, the codes are essentially comparable with the Golay code being slightly superior in performance, while the convolutional code may be simpler to implement in hardware.

The implementation of the code is straight forward, the encoder is shown in Figure 3.7 and the decoder in Figure 3.8.

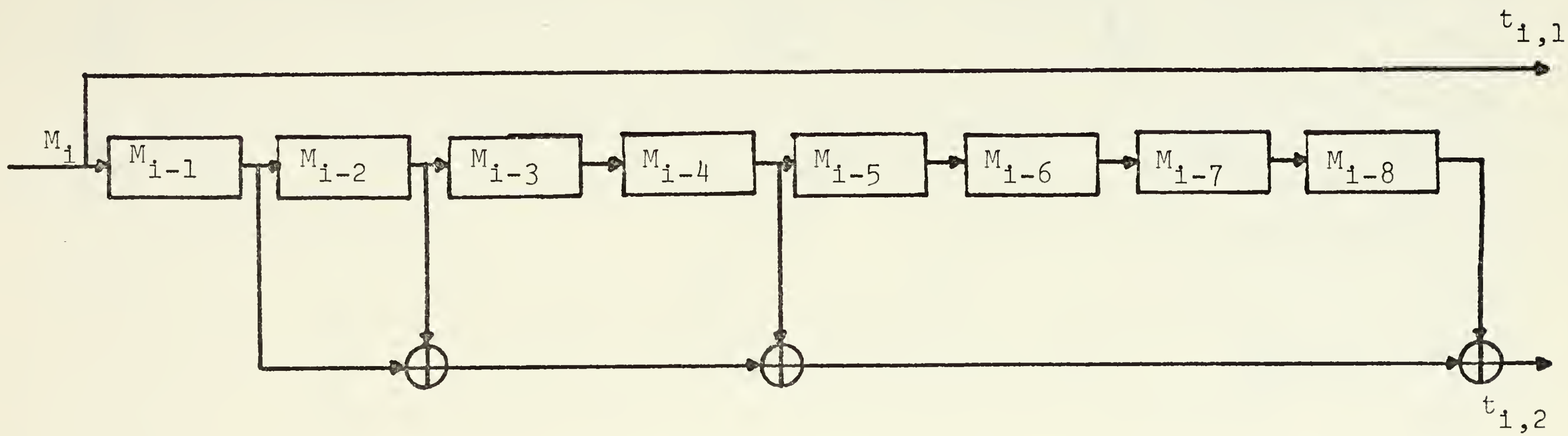

Figure 3.7. Encoder for proposed rate ½ convolutional code

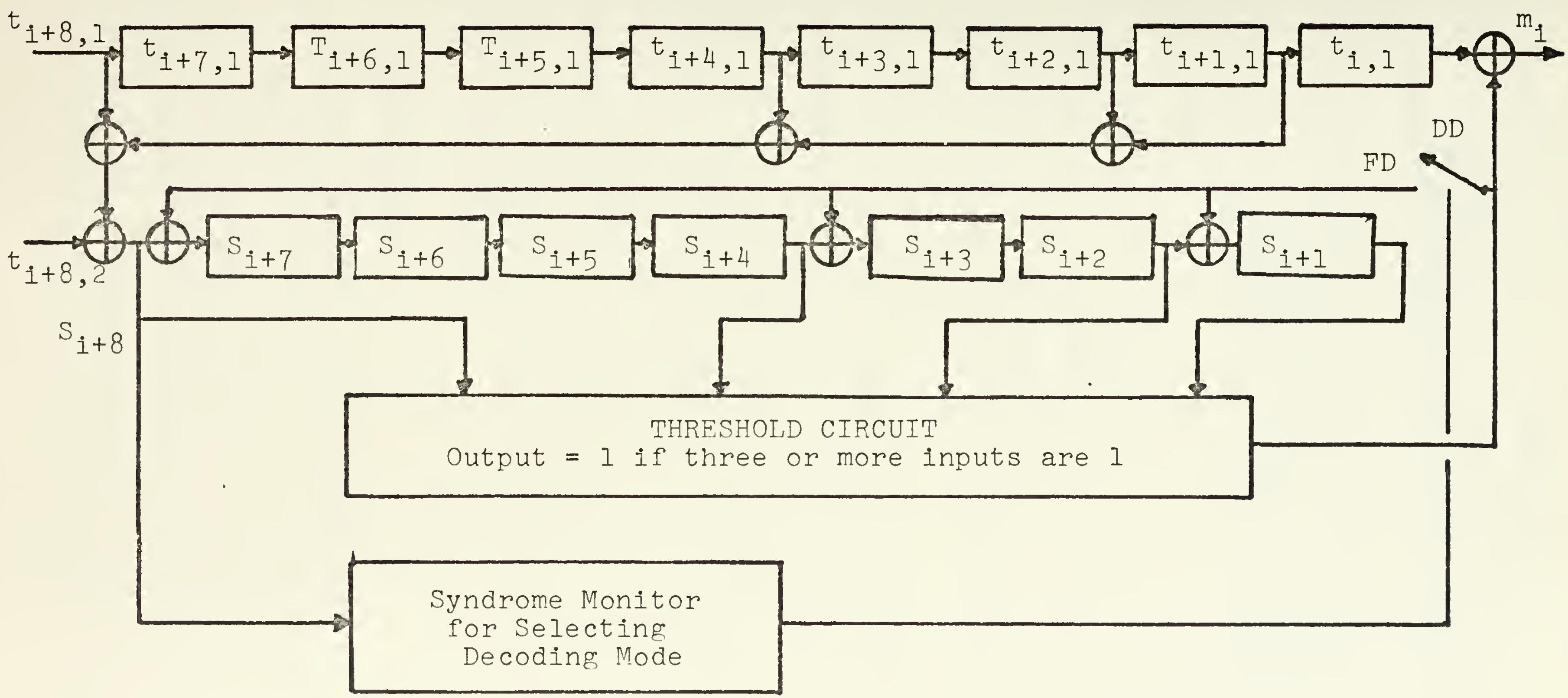

Figure 3.8. Decoder for code generated by encoder of Figure 3.7

TABLE III

A COMPARISON OF BIT ERROR PROBABILITIES FOR CONVOLUTIONAL CODE AND MODIFIED GOLAY CODE, CHANNEL ERROR PROBABILITY, IS APPROXIMATELY 2×10^{-2}

NI	Decoded Bit Error Probability Convolutional Code	Decoded Bit Error Probability Modified Golay Code
480	1.5×10^{-3}	5.3×10^{-4}
960	1.2×10^{-3}	3.4×10^{-4}
1920	7.86×10^{-4}	2.3×10^{-4}
15360	5.9×10^{-4}	1.1×10^{-4}

In this case, each rectangular box represents a 54 stage shift register. The mode selector examines the general state of the syndrome shift registers and if there are numerous indicated errors the mode is changed from feedback decoding to definite decoding, the decoder remains in the definite decoding mode until such time as there are relatively few indicated errors when the decoder is returned to the feedback mode.

F. CONCLUSIONS

The improvement of the reliability of the HF communication channel through the use of error correction coding has been the major concern of this thesis. The average bit error probability is determined by the number and duration of fades, and atmospheric noise. The fading is assumed to be an interference phenomena resulting from the existence

of multiple propagation paths, and lasting for periods of approximately one second. The atmospheric noise contributes random errors to the normally burst characteristics produced by the fades.

It is demonstrated that simple convolutional codes utilized with moderate amounts of interleaving can obtain approximately an order of magnitude reduction in the output bit error probability. In addition, the relatively simple decoding scheme may be modified during reception to eliminate the possibility of error propagation, with only a relatively minor loss in the error correction capability of the code.

In order to fully utilize the channel other procedures should be followed, namely space diversity, where two or more separate receivers are utilized with antennas in physically different locations. This decreases the effect of fading resulting from the multipath conditions since the fading at different locations is generally uncorrelated. In addition, frequency diversity (when possible) should be employed, since the fades are unrelated for carrier frequencies separated by more than 50 KHz.

Since the normal allocated bandwidth in the HF spectrum is 3 KHz, many separate channels may be frequency division multiplexed onto the same carrier to achieve a higher data rate on the channel. Existing frequency division multiplex equipments have a 32 channel capability, with the sub-channels each operating at a 75 bit per second rate for a maximum data rate of 2400 bits per second. The error

correction coding will require a sacrifice in the data rate on the channel to provide a reduction in the output error rate. By using a rate ½ code the information rate through the channel will be reduced to 1200 bits per second, 16 subchannels may then be used for the information bits and the remaining 16 subchannels for the parity bits. One method of implementing such a system indicated in Figure 3.9.

G. RECOMMENDATIONS FOR FUTURE RESEARCH

There are three primary areas where considerable research could be applied: these are 1) the characterization of the HF channel, 2) the modeling of man-made noise and interference, and 3) the decoding of convolutional codes.

In the characterization of the HF channel the use of the Markov chain model is most promising, however, the principal source of difficulty is the non-availability of data. The data analyzed has been obtained from actual channel measurements which may have been perturbed by categroically diffrent propagation conditions. To avoid these interpretations, it is recommended that existing ionospheric channel simulators (24) be used to obtain characteristics which are dependent only on multipath phenomena. Actual channel data could then be compared with that from the simulator to determine other sources of perturbation.

In the area of impulsive man-made noise or signal interference maximum use could be made of the Hall model. The

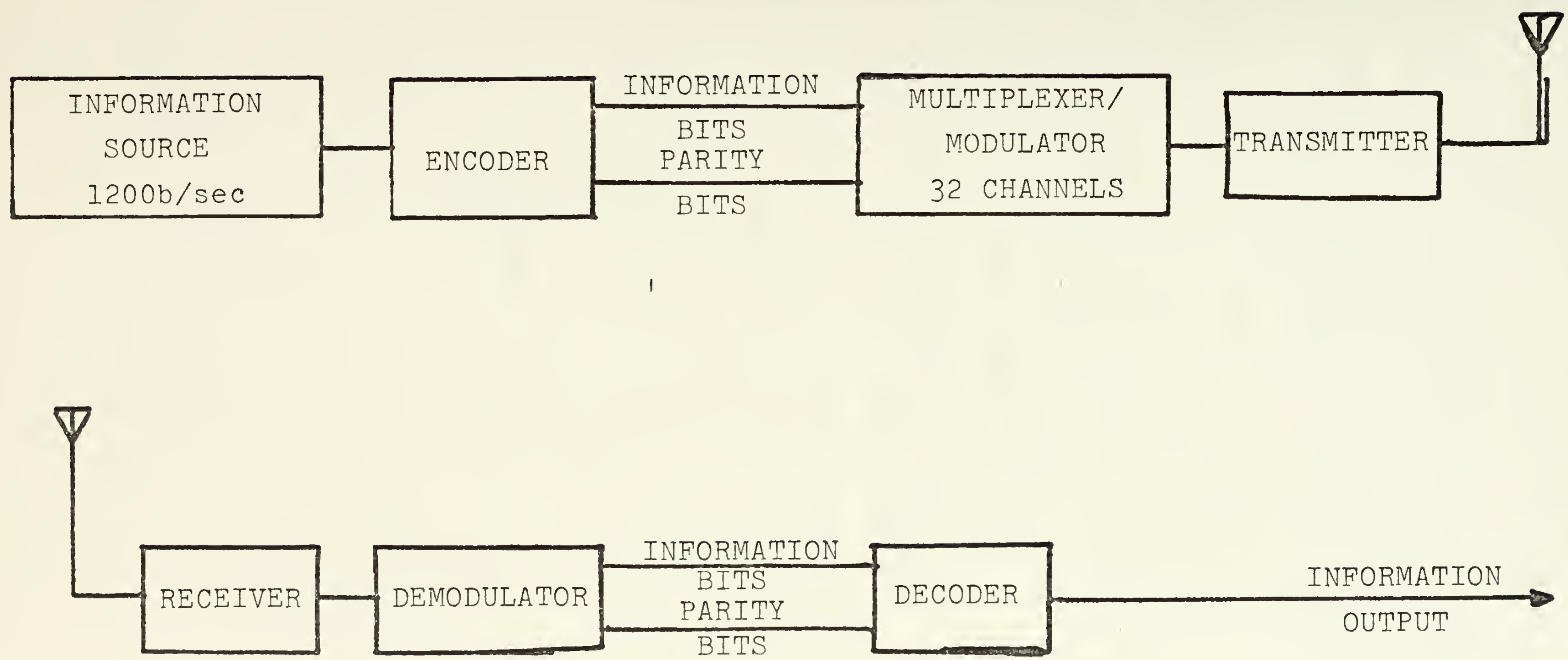

Figure 3.9. Implementation of 1200 bit per second system

Hall model was discussed in this thesis only for the amplitude characteristics of atmospheric noise. However, its principal advantage lies in the fact that statistics such as pulse interval and pulse duration may be correctly obtained by proper specification of the covariance function of the modulating process, A(t), as described on page 28. This would be a valuable tool in the design of receivers for use in areas where unavoidable noise sources (such as powerline transients) cause excessive error rates with conventional receivers.

One of the most important fields for additional research is in the area of decoding algorithms for convolutional codes. In general, there are some codes with desirable distance parameters, but good general decoding algorithms are not available. Threshold decoding, while relatively easy to implement, is not applicable to all codes. Sequential decoding, on the other hand, will work with any convolutional code, but it will not work well on channels where there is correlation between errors, i.e., burst channels. While the Viterbi decoding algorithm is a maximum likelihood decoding procedure, it is limited to codes of relatively short constraint length by the storage requirements. In short, there are no known classes of convolutional codes with decoding procedures comparable with the longer Bose-Chaudhuri-Hocquenghem block codes.

LIST OF REFERENCES

1. Davies, K., Ionospheric Radio Propagation, p. 159-214, National Bureau of Standards, Monograph 80, April 1965.

2. Institutes for Environmental Research IER 44- ITSA 44, Signal Amplitude Distribution for Stationary Radio Multipath Conditions by M. Nesenbergs, p. 67-68, 1-12, 55, 82, Aug. 1967.

3. Abramowitz, M., and Stegun, I. A., Handbook of Mathematical Functions, U.S. Department of Commerce, AMS 55.

4. National Bureau of Standards Report 22.4, A Comparison of Amplitude and Angle Modulation for Narrow Band Communication of Binary-Coded Messages in Fluctuating Noise, by G. F. Montgomery, p. 3-12, 10, 11 March 1953.

5. Tsai, S., "Markov Characterization of the HF Channel," IEEE Transactions on Communication Technology, Vol COM-17 No. 1, p. 24-32, February 1969.

6. McManamon, P., "HF Markov Chain Models and Measured Error Averages," IEEE Transactions on Communication Technology, Vol. COM-18, No. 3, p. 2-1-208, June 1970.

7. International Radio Consultative Committee, CCIR Report 322, World Distribution and Characteristics of Atmospheric Radie Noise, p. 4-16, International Telecommunication Union, 1964.

8. Bello, P. A., Esposito, R., "A New Method for Calculating Probabilities of Errors Due to Impulsive Noise," IEEE Transactions on Communication Technology, Vol, COM-17, No. 3, June 1969.

9. Engel, J. S., "Digital Transmission in the Presence of Impulsive Noise," Bell System Technical Journal, Vol. 44, p. 1699-1743, October 1965.

10. Stanford Electronics Laboratory Report SU-SEL-66-052, A New Model for "Impulsive" Phenomena: Application to Atmospheric-Noise Communication Channels, by H. M. Hall, p. 11-24, Aug. 1966.

11. Parzan, E., Modern Probability Theory and Its Applications, p. 318, Wiley, 1960.

12. ESSA Technical Memorandum ERLTM-ITS 184, SAMSO Phase C-Final Report, Noise Data and Analysis, by A. D. Spaulding, R. T. Disney, and L. R. Espeland, p. 13, June 1969.

13. Fritchman, B. D., "A Binary Channel Characterization Using Partitioned Markov Chains," IEEE Transactions on Information Theory, Vol IT-13, No. 2, p. 221-227, April 1967.

14. Cox, D. R. and Miller, H. D., The Theory of Stochastic Processes, p. 95, Spottiswode, Ballantyne and Co. 1965.

15. Massey, J. M., Threshold Decoding, MIT Press, 1963.

16. Wyner, A. D., and Ash, R. B., "Analysis of Recurrent Codes," IEEE Transactions on Information Theory, Vol. IR-9, p. 143-156, July 1963.

17. Robinson, J. P., and Bernstein, A. J., "A Class of Binary Recurrent Codes with Limited Error Propagation," IEEE Transactions on Information Theory, Vol. IT-13. No. 1, p. 106-113, January 1967.

18. Robinson, J. P., "Error Propagation and Definite Decoding of Convolutional Codes," IEEE Transactions on Information Theory, Vol. IT-14, No. 1, p. 121-128, January 1968.

19. Wozencraft, J. M., "Sequantial Decoding for Reliable Communication," IRE Nat. Conv. Re. Vol. 5. Pt, p. 11-25, 1957.

20. Lin, S., and Lyne, H., "Some Results on Binary Convolutional Code Generators," IEEE Transactions on Information Theory, Vol. IT-13, No. 1, p. 134-139, January 1967.

21. Viterbi, A. J., "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," IEEE Transactions on Information Theory, Vol. IT-13, p. 260-269, 1967.

22. Omura, J. K., "On the Viterbi Decoding Algorithm," IEEE Transactions on Information Theory, Vol. IT-15, p. 177-179, January 1969.

23. Systems Research Corporation, EDAC Coding for Naval HF Communications, by W. W. Peterson, N. Abramson, N. T. Gaardner, and C. L. Chen, p. I-20-I-22, May 1969.

24. ESSA Technical Report ERL-112-ITS 80, Experimental Verification of an Ionospheric Channel Model, by C. C. Watterson, J. R. Juroshek, and W. D. Bensema, July 1969.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Professor G. H. Marmont Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	1
4.	LT James B. McCrumb Chief of Naval Materiel MAT 03423 Navy Department Washington, D. C. 20360	1

Blank

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

THE APPLICATION OF CONVOLUTIONAL CODES TO THE HIGH FREQUENCY CHANNEL

4. DESCRIPTIVE NOTES *(Type of report and, inclusive dates)*

Electrical Engineer's Thesis; September 1970

5. AUTHOR(S) *(First name, middle initial, last name)*

James Brayton McCrumb, Lieutenant, United States Navy

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1970	97	24

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

The HF channel is characterized with regard to frequency selective fading and atmospheric noise under restrictive mathematical assumptions, and compared with measured statistics. Actual channel characteristics are utilized to obtain parameters for choosing appropriate convolutional codes. Short threshold decodable convolutional codes with only a moderate amount of interleaving are shown to be capable of affecting a reduction in the output error probability by at least an order of magnitude, for uncoded error probabilities on the order of 2×10^{-2}.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
HF COMMUNICATION CHANNEL						
FREQUENCY SELECTIVE FADING						
ATMOSPHERIC NOISE						
CONVOLUTIONAL CODE						
THRESHOLD DECODING						

Thesis 122997
M1826 McCrumb
c.1 The application of
convolutional codes
to the high frequency
channel.